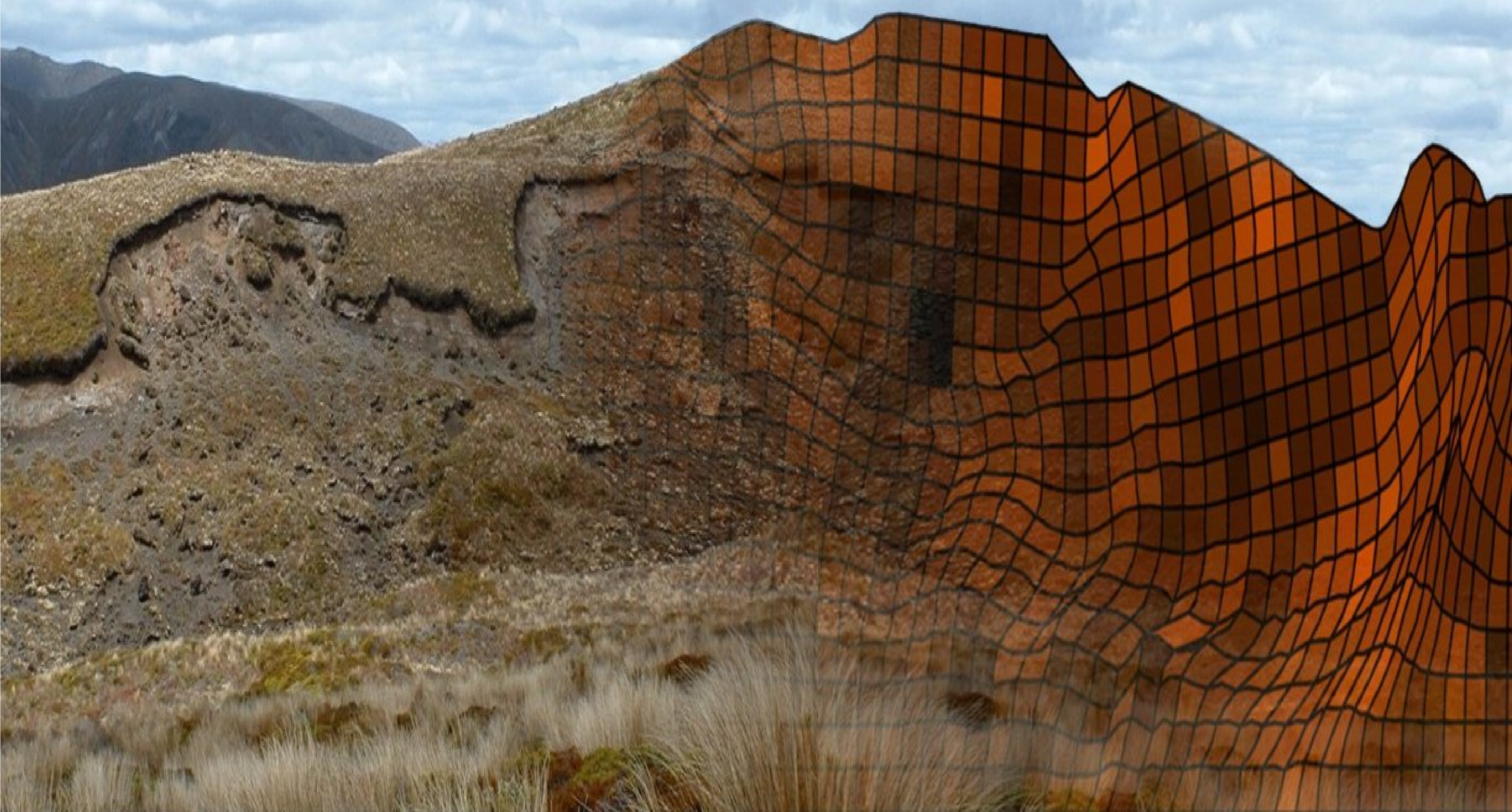
Common Distributions

in *Stochastic Analysis and Inverse Modeling*

Presented by

Gordon A. Fenton



Common Discrete Distributions

Bernoulli Trials:

- form the basis for 6 very common distributions (binomial, geometric, negative binomial, Poisson, exponential, and gamma)

Fundamental Assumptions;

1. Each trial has two possible outcomes (0/1, true/false, success/failure, on/off, yes/no, etc)
2. Trials are independent (allows us to easily compute probabilities)
3. Probability of "success" remains constant from trial to trial

Note: if we relax assumptions 1 and 2 we get Markov Chains, another suite of powerful stochastic models

Bernoulli Trials

Let $X_i = 1$ if the i'th trial is a "success"

$= 0$ if not

then the Bernoulli probability distribution is

$$P[X_i = 1] = p$$

$$P[X_i = 0] = 1 - p = q$$

Bernoulli Trials

$$\mathrm{E}\left[X_i\right] = \sum_{j=0}^{1} j\,\mathrm{P}\left[X_i = j\right] = 0(q) + 1(p) = p$$

$$\mathrm{E}\left[X_i^{\,2}\right] = \sum_{j=0}^{1} j^2\,\mathrm{P}\left[X_i = j\right] = 0^2(q) + 1^2(p) = p$$

$$\mathrm{Var}\left[X_i\right] = \mathrm{E}\left[X_i^2\right] - \mathrm{E}^2\left[X_i\right] = p - p^2 = pq$$

Common Discrete Distributions

Bernoulli Family of Distributions

Discrete Trials	Every "Instant" Becomes a Trial
1) Binomial: N_n = number of "successes" in n trials	4) Poisson: N_t = number of "successes" in time t
2) Geometric: T_1 = number of trials until the first "success"	5) Exponential: (continuous) T_1 = time to the first "success"
3) Negative Binomial: T_k = number of trials until the k'th "success"	6) Gamma: (continuous) T_k = time to the k'th "success"

Binomial Distribution

If $N_n =$ number of "successes" in n trials, then

$$N_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

since $X_i = 1$ if the trial results in a "success" and 0 if not.

Mean: $$\mathrm{E}[N_n] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} p = np$$

Variance:

$$\mathrm{Var}[N_n] = \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] \ \text{(since independent)}$$

$$= \sum_{i=1}^{n} pq = npq$$

Binomial Distribution

Consider $P[N_5 = 2] = P[\{SSFFF\} \cup \{SFSFF\} \cup \cdots \cup \{FFFSS\}]$

$$= P[SSFFF] + P[SFSFF] + \cdots + P[FFFSS]$$

$$= P[S]P[S]P[F]P[F]P[F] + P[S]P[F]P[S]P[F]P[F] + \cdots$$

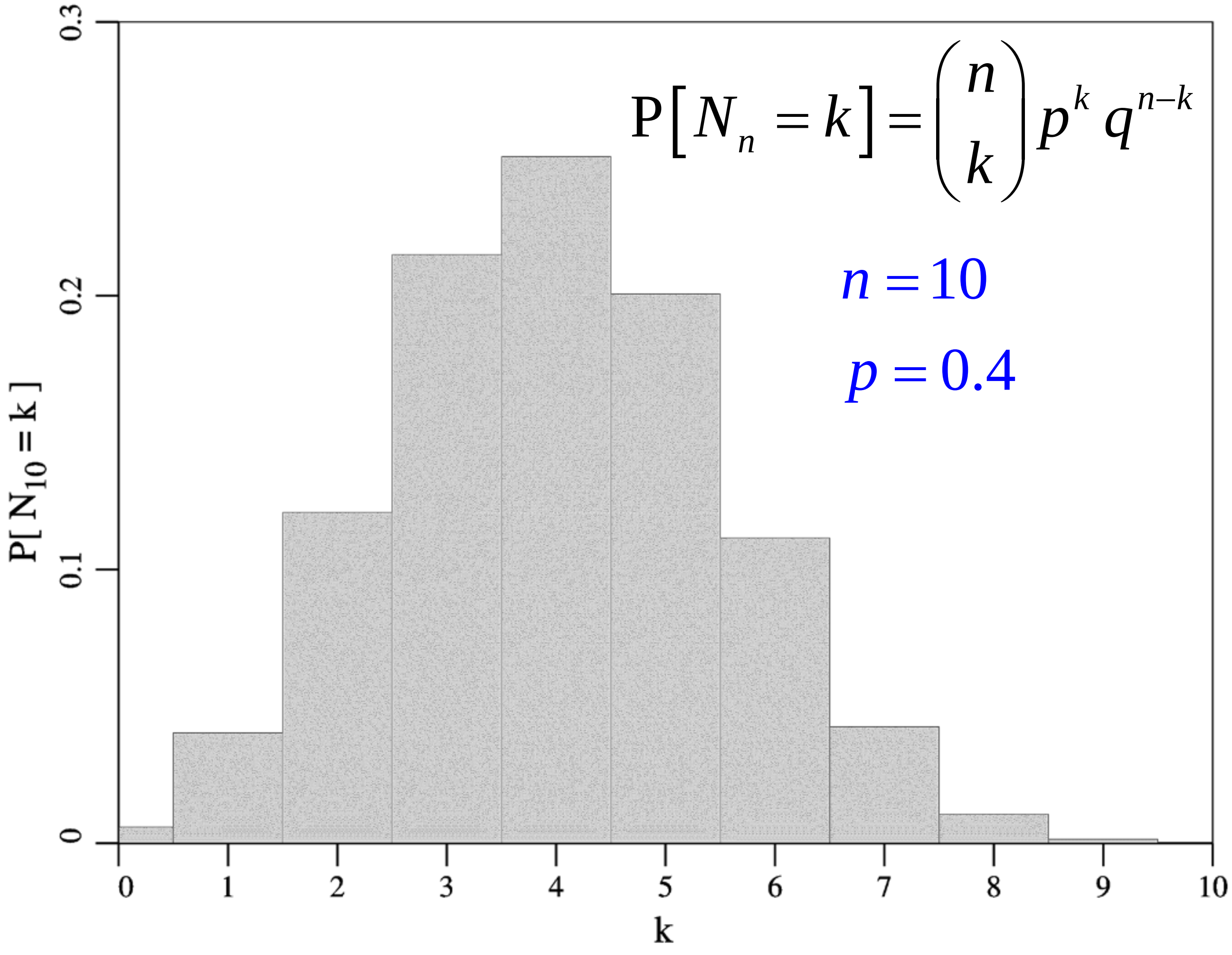
$$= p^2q^3 + p^2q^3 + \cdots + p^2q^3$$

$$= \binom{5}{2} p^2 q^3$$

In general the binomial distribution is

$$P[N_n = k] = \binom{n}{k} p^k q^{n-k}$$

Binomial Distribution



Binomial Distribution: Example

Suppose that the probability of failure, p_f, of a slope is investigated using Monte Carlo simulation. That is, a series of $n = 5000$ independent realizations of the slope strength are simulated from a given probability distribution and the number of failures counted to be $n_f = 120$. What is the estimated failure probability and what is the 'standard error' (standard deviation) of this estimate?

Solution: each realization is an independent trial with two possible outcomes, success or failure, having constant probability of failure (p_f). Thus, each realization is a Bernoulli trial, and the number of failures is a realization of N_n, which follows a binomial distribution.

Binomial Distribution: Example

We have $n = 5000$ and $n_f = 120$

The failure probability is estimated by

$$\hat{p}_f = \frac{\text{number of failures}}{\text{number of trials}} = \frac{120}{5000} = 0.024$$

In general $\hat{p}_f = \dfrac{N_n}{n}$

so that $\sigma_{\hat{p}_f}^2 = \mathrm{Var}\left[\hat{p}_f\right] = \mathrm{Var}\left[\dfrac{N_n}{n}\right] = \left(\dfrac{1}{n^2}\right)\mathrm{Var}[n]$

$$= \frac{npq}{n^2} = \frac{pq}{n}$$

which, for $p \simeq \hat{p}_f = 0.024$ gives us

$$\sigma_{\hat{p}_f} \simeq \sqrt{\frac{0.024(1-0.024)}{5000}} = 0.0022$$

Geometric Distribution

Let $T_1 =$ number of trials until the first "success"

Consider $\mathrm{P}\left[T_1 = 4\right] = \mathrm{P}\left[FFFS\right] = pq^3$

In general $\mathrm{P}\left[T_1 = k\right] = pq^{k-1}$

$$\mathrm{E}\left[T_1\right] = \sum_{k=1}^{\infty} kpq^{k-1} = \frac{1}{p}$$

$$\mathrm{Var}\left[T_1\right] = \mathrm{E}\left[T_1^2\right] - \mathrm{E}^2\left[T_1\right] = \sum_{k=1}^{\infty} k^2 pq^{k-1} - \frac{1}{p^2} = \frac{q}{p^2}$$

Geometric Distribution

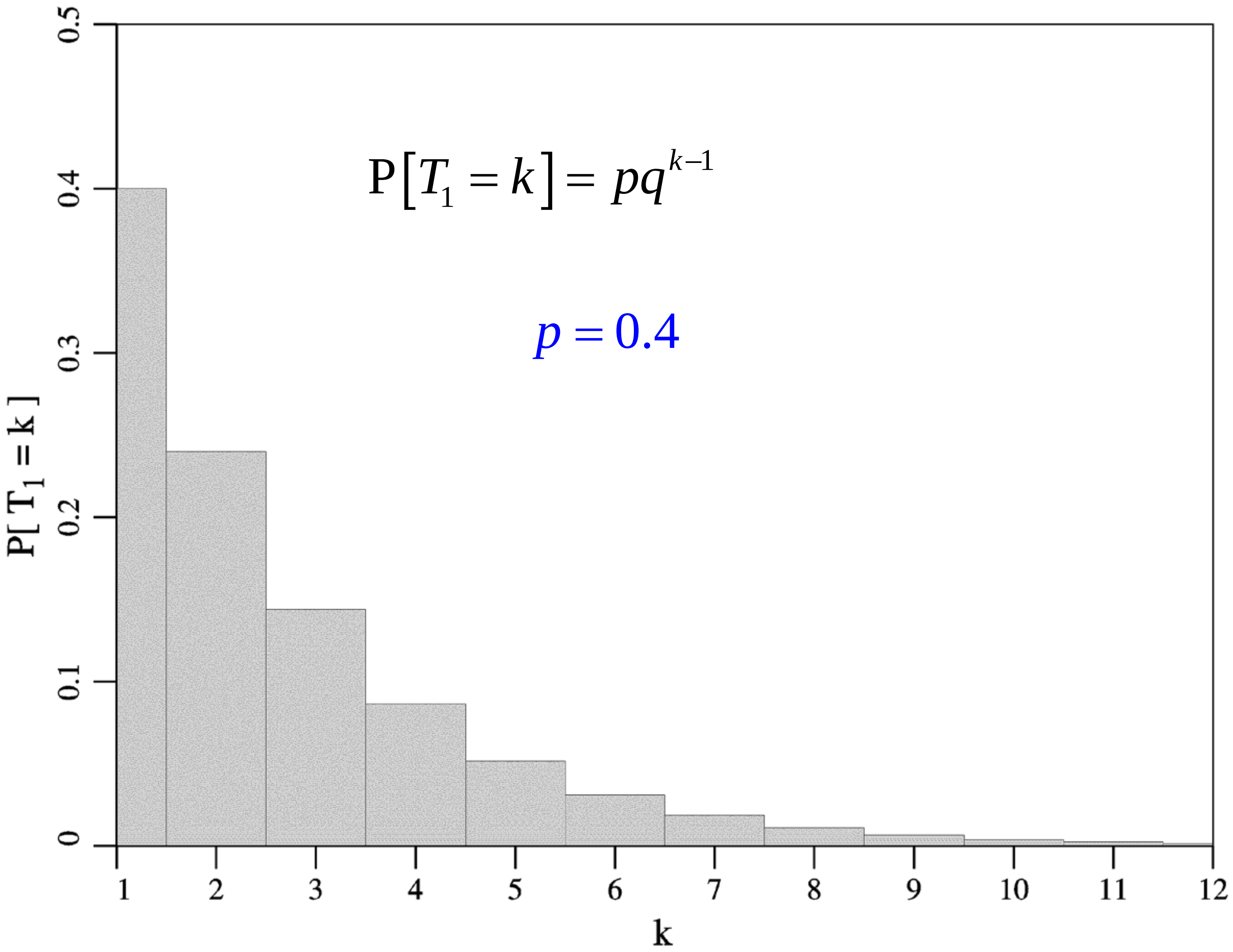
Because all trials are independent, the geometric distribution is "memoryless". That is, it doesn't matter when we start looking at a sequence of trials – the distribution of the number of trials to the next "success" remains the same. That is,

$$\mathrm{P}\left[T_1 > j+k \mid T_1 > j\right] = \frac{\mathrm{P}\left[T_1 > j+k\right]}{\mathrm{P}\left[T_1 > j\right]} = \frac{\sum_{m=j+k+1}^{\infty} pq^{m-1}}{\sum_{n=j+1}^{\infty} pq^{n-1}} = q^k$$

and

$$\mathrm{P}\left[T_1 > k\right] = \sum_{m=k+1}^{\infty} pq^{m-1} = p\sum_{m=k}^{\infty} q^m = pq^k \sum_{m=0}^{\infty} q^m = pq^k\left(\frac{1}{1-q}\right) = q^k$$

Geometric Distribution



Geometric Distribution: Example

A series of piles have been designed to be able to withstand a certain test load with probability $p = 0.8$. If the resulting piles are sequentially tested at that design load, what is the probability that the first pile to fail the test is the 7'th pile tested?

Solution: If piles can be assumed to fail independently with constant probability, then this is again a sequence of Bernoulli trials. We thus want to compute

$$\mathrm{P}\left[T_1 = 7\right] = pq^{7-1} = (0.8)(0.2)^6 = 0.00005$$

Negative Binomial Distribution

Let T_k = the number of trials until the k'th "success"

Consider

$$P\left[T_3 = 5\right] = P\left[\{SSFFS\} \cup \{SFSFS\} \cup \cdots \cup \{FFSSS\}\right]$$

$$= P\left[SSFFS\right] + P\left[SFSFS\right] + \cdots + P\left[FFSSS\right]$$

$$= \binom{4}{2} p^3 q^2$$

In general $P\left[T_k = m\right] = \binom{m-1}{k-1} p^k q^{m-k}$

Negative Binomial Distribution

The negative binomial distribution is a generalization of the geometric distribution (the *k*'th "success"). Its mean and variance can be obtained by realizing that T_k is the sum of *k* independent T_1's. That is

$$T_k = T_{1_1} + T_{1_2} + \cdots + T_{1_k}$$

so that

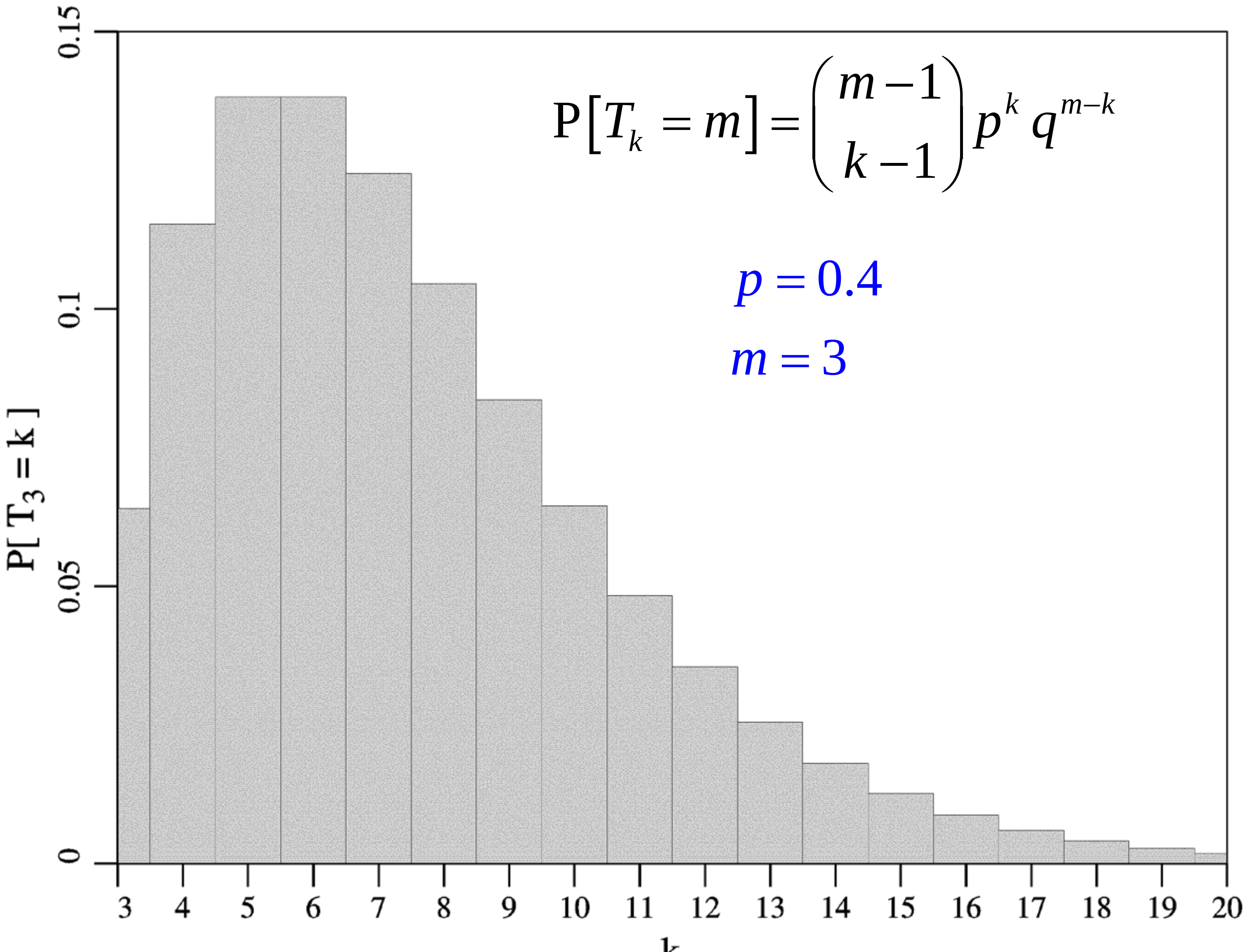
$$\mathrm{E}[T_k] = \mathrm{E}\left[\sum_{j=1}^{k} T_{1_j}\right] = \sum_{j=1}^{k} \mathrm{E}[T_1] = \frac{k}{p}$$

and

$$\mathrm{Var}[T_k] = \mathrm{Var}\left[\sum_{j=1}^{k} T_{1_j}\right] = \sum_{j=1}^{k} \mathrm{Var}[T_1] \quad \text{(since independent)}$$

$$= k\mathrm{Var}[T_1] = \frac{kq}{p^2}$$

Negative Binomial Distribution



Negative Binomial Distribution: Example

A series of piles have been designed to be able to withstand a certain test load with probability $p = 0.8$. If the resulting piles are sequentially tested at that design load, what is the probability that the third pile to fail the test is the 7'th pile tested?

Solution: If piles can be assumed to fail independently with constant probability, then this is again a sequence of Bernoulli trials. We thus want to compute

$$\mathrm{P}[T_3 = 7] = \binom{k-1}{m-1} p^m q^{k-m} = \binom{7-1}{3-1} (0.8)^3 (0.2)^{7-3}$$

$$= \binom{6}{2} (0.8)^3 (0.2)^4 = 0.00082$$

Poisson Distribution

We now let every instant in time (or space) become a Bernoulli trial (i.e., a sequence of independent trials)

“Successes” can now occur at any instant in time.

PROBLEM: in any time interval, there are now an infinite number of Bernoulli trials. If the probability of “success” of a trial is, say, 0.2, then in any time interval we get an infinite number of “successes”. This is not a very useful model!
SOLUTION: we can no longer talk about probability of “success” of individual trials, p (which becomes zero). We must now talk about the mean *rate* of “success”, λ.

Poisson Distribution

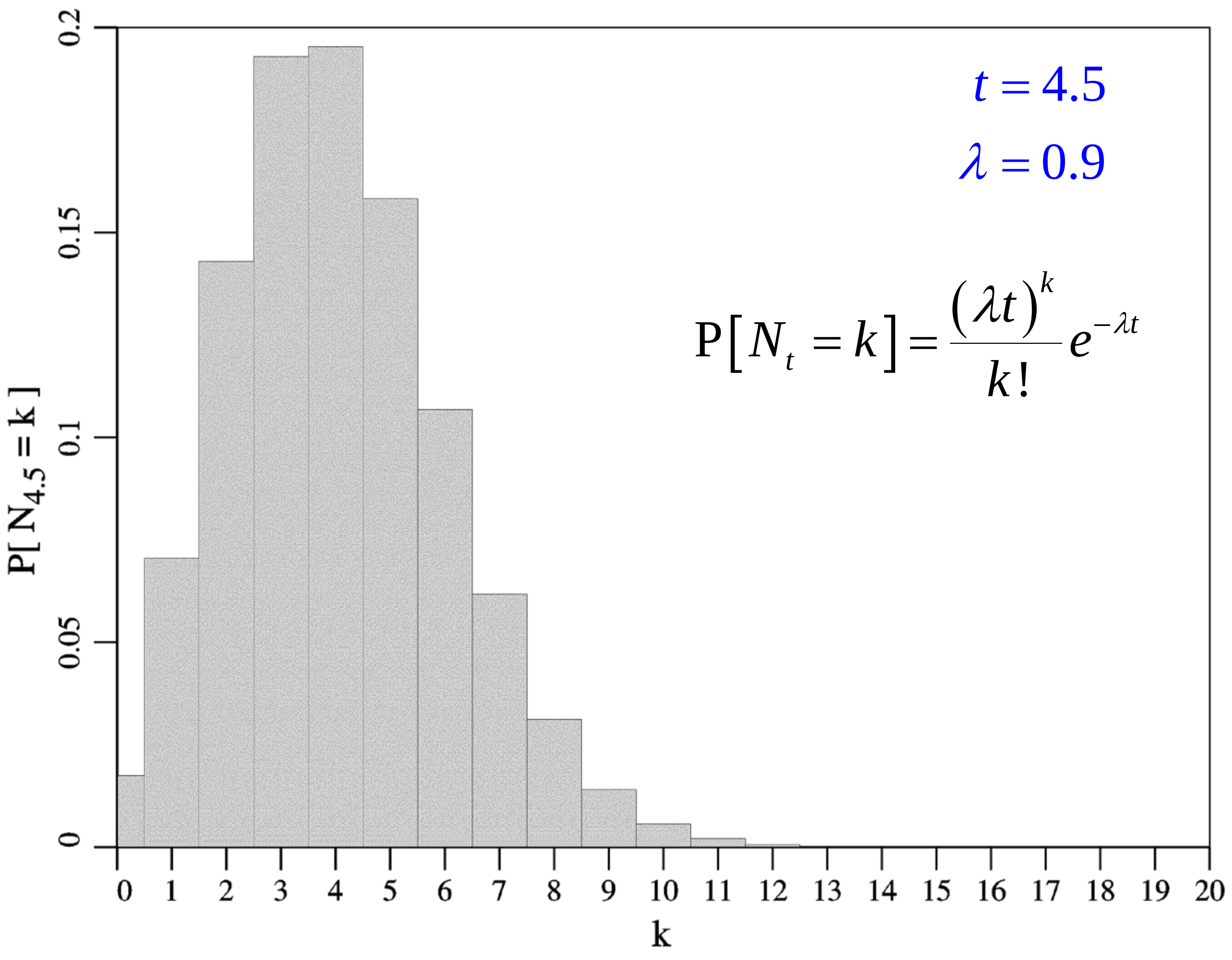
Governs many rate dependent processes (arrivals of vehicles at an intersection, 'packets' through an internet gateway, earthquakes exceeding magnitude M, floods, load extremes, etc.)

Let N_t = number of "successes" (arrivals) in time interval t

The distribution of N_t is arrived at in the limit as the number of trials goes to infinity assuming the probability of "success" is proportional to the number of trials (see the course notes for details). We get

$$\mathrm{P}\left[N_t = k\right] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Poisson Distribution



Poisson Distribution: Example

Many research papers suggest that the arrivals of earthquakes follow a Poisson process over time. Suppose that the mean time between earthquakes is 50 years at a particular location.

1. How long must the time period be so that the probability that no earthquakes occur during that period is at most 0.1?

Solution: $\lambda = 1/\mathrm{E}\left[T_1\right] = 1/50 = 0.02$

1. Let N_t be the number of earthquakes over the time interval t. We want to find t such that

$$\mathrm{P}\left[N_t = 0\right] = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} \le 0.1$$

this gives $t \ge -\ln(0.1)/\lambda = -\ln(0.1)/0.02 = 115$ years

Poisson Distribution: Example

2. Suppose that 50 years pass without any earthquakes occurring. What is the probability that another 50 years will pass without any earthquakes occurring?

Solution:

$$P[N_{100}=0 \mid N_{50}=0] = \frac{P[N_{100}=0 \cap N_{50}=0]}{P[N_{50}=0]} = \frac{P[N_{100}=0]}{P[N_{50}=0]}$$

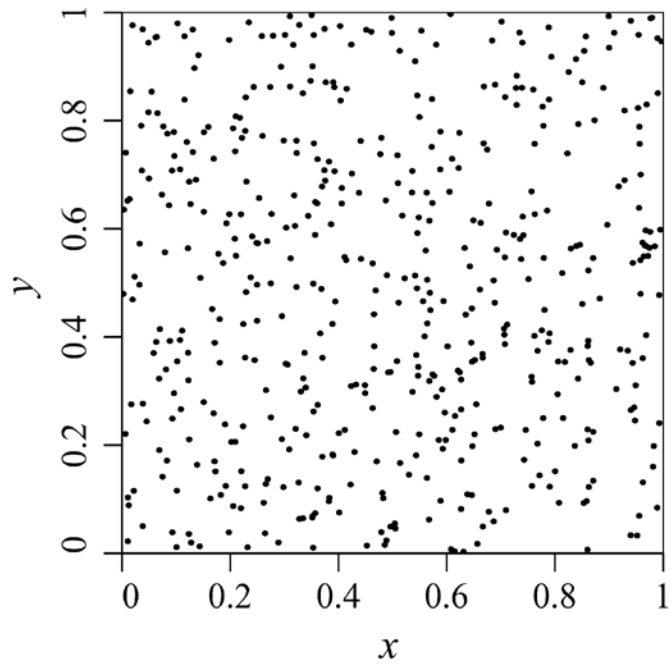
$$= \frac{e^{-100\lambda}}{e^{-50\lambda}} = e^{-50\lambda} = e^{-1} = 0.368$$

Note that due to memorylessness,

$$P[N_{50}=0] = e^{-50\lambda} = e^{-1} = 0.368$$

Simulating a Poisson Distribution

Since the Poisson process derives from an infinite number of independent and equilikely Bernoulli trials over time (or space), its simulation is simple: uniformly distributed (equilikely) on any interval.



Common Continuous Distributions

The Bernoulli Family extends to two continuous distributions when every instant in time (space) becomes a Bernoulli trial;

1. Exponential: T_1 = time until the next "success" (arrival). This is the continuous time analog to the geometric distribution.

2. Gamma: T_k = time until the k'th success (or arrival). This is the continuous time analog to the negative binomial distribution.

Exponential Distribution

Consider the Poisson Distribution which governs the number of "successes" in time interval t when every instant is an independent Bernoulli trial. We know that the Poisson distribution specifies that

$$\mathrm{P}\left[N_t = k\right] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

What is the probability that the time to the first "success" is greater than *t*? If the first "success" occurs later than *t*, then the number of successes in time interval *t*, must be zero, i.e.

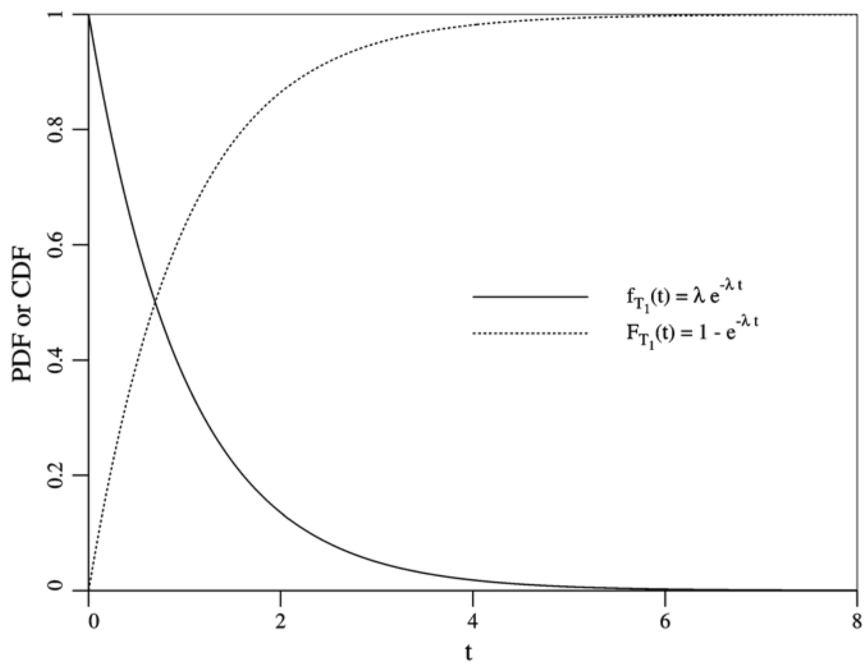
$$\mathrm{P}\left[T_1 > t\right] = \mathrm{P}\left[N_t = 0\right] = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

Exponential Distribution

If $\mathrm{P}[T_1 > t] = \mathrm{P}[N_t = 0] = \dfrac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$

then $F_{T_1}(t) = \mathrm{P}[T_1 \le 1] = 1 - e^{-\lambda t}$

so that $f_{T_1}(t) = \dfrac{dF_{T_1}(t)}{dt} = \lambda e^{-\lambda t}, \;\; t \ge 0$



$$\mathrm{E}[T_1] = \frac{1}{\lambda}$$

$$\mathrm{Var}[T_1] = \frac{1}{\lambda^2}$$

Exponential Distribution

Again, by virtue of the fact that every instant in time (space) is an independent Bernoulli trial, the exponential distribution is also memoryless. That is,

$$\mathrm{P}\left[T_1 > t+s \mid T_1 > t\right] = \frac{\mathrm{P}\left[T_1 > t+s \cap T_1 > t\right]}{\mathrm{P}\left[T_1 > t\right]} = \frac{\mathrm{P}\left[T_1 > t+s\right]}{\mathrm{P}\left[T_1 > t\right]}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s}$$

but since $\mathrm{P}\left[T_1 > s\right] = e^{-\lambda s}$ it is apparent that it doesn't matter when you start looking. I.e., the probability that the next "success" is time *s* in the future is independent of *t* (when we start looking).

Exponential Distribution

The "memoryless" property of the exponential distribution means that it is a common "lifetime" model (where "success" is defined as "failure"!) in systems that neither improve nor degrade with time – failure probability remains constant with time.

However, some systems improve with time (e.g. early strength of concrete) and some systems degrade with time (e.g. fatigue), which leads to the Weibull distribution to be discussed shortly.

Exponential Distribution: Example

Let us assume that earthquakes in a certain region occur on average once every 50 years and that the number of earthquakes in any time interval follows a Poisson distribution. Under these conditions, what is the probability that less than 30 years will pass before the next earthquake occurs?

Solution: Let T_1 be the time to the next earthquake. Then, since the number of earthquakes follow a Poisson distribution, the time between earthquakes follows an exponential distribution. Thus, T_1 follows an exponential distribution with $\lambda = 1/50 = 0.02$ earthquakes per year (on average), and

$$\mathrm{P}\left[T_1 < 30 \text{ years}\right] = 1 - e^{-30(0.02)} = 0.549$$

Gamma Distribution

The last of the Bernoulli Family of distributions is a generalization of the exponential distribution.

Let $T_k =$ time until the k'th "success"

which is evidently the sum of the times between the first k "successes", i.e.

$$T_k = T_{1_1} + T_{1_2} + \cdots + T_{1_k}$$

which leads to the following moments

$$\mathrm{E}\left[T_k\right] = k\,\mathrm{E}\left[T_1\right] = \frac{k}{\lambda}$$

$$\mathrm{Var}\left[T_k\right] = k\,\mathrm{Var}\left[T_1\right] = \frac{k}{\lambda^2}$$

Gamma Distribution

To determine the Gamma distribution, lets consider an example: $\mathrm{P}\left[T_3 > t\right]$ means the probability that the 3rd "success" occurs after time *t*.

For this to be true, we must have had no more than 2 "successes" within time interval *t*. That is, the event $T_3 > t$ is equivalent to the event $N_t \le 2$

In other words:

$$\mathrm{P}\left[T_3 > t\right] = \mathrm{P}\left[N_t \le 2\right] = \mathrm{P}\left[N_t = 0\right] + \mathrm{P}\left[N_t = 1\right] + \mathrm{P}\left[N_t = 2\right]$$

$$= e^{-\lambda t} + \frac{(\lambda t)^1}{1!} e^{-\lambda t} + \frac{(\lambda t)^2}{2!} e^{-\lambda t}$$

$$= e^{-\lambda t} \sum_{j=0}^{2} \frac{(\lambda t)^j}{j!}$$

Gamma Distribution

If $\mathrm{P}\left[T_3 > t\right] = e^{-\lambda t} \sum_{j=0}^{2} \frac{(\lambda t)^j}{j!}$

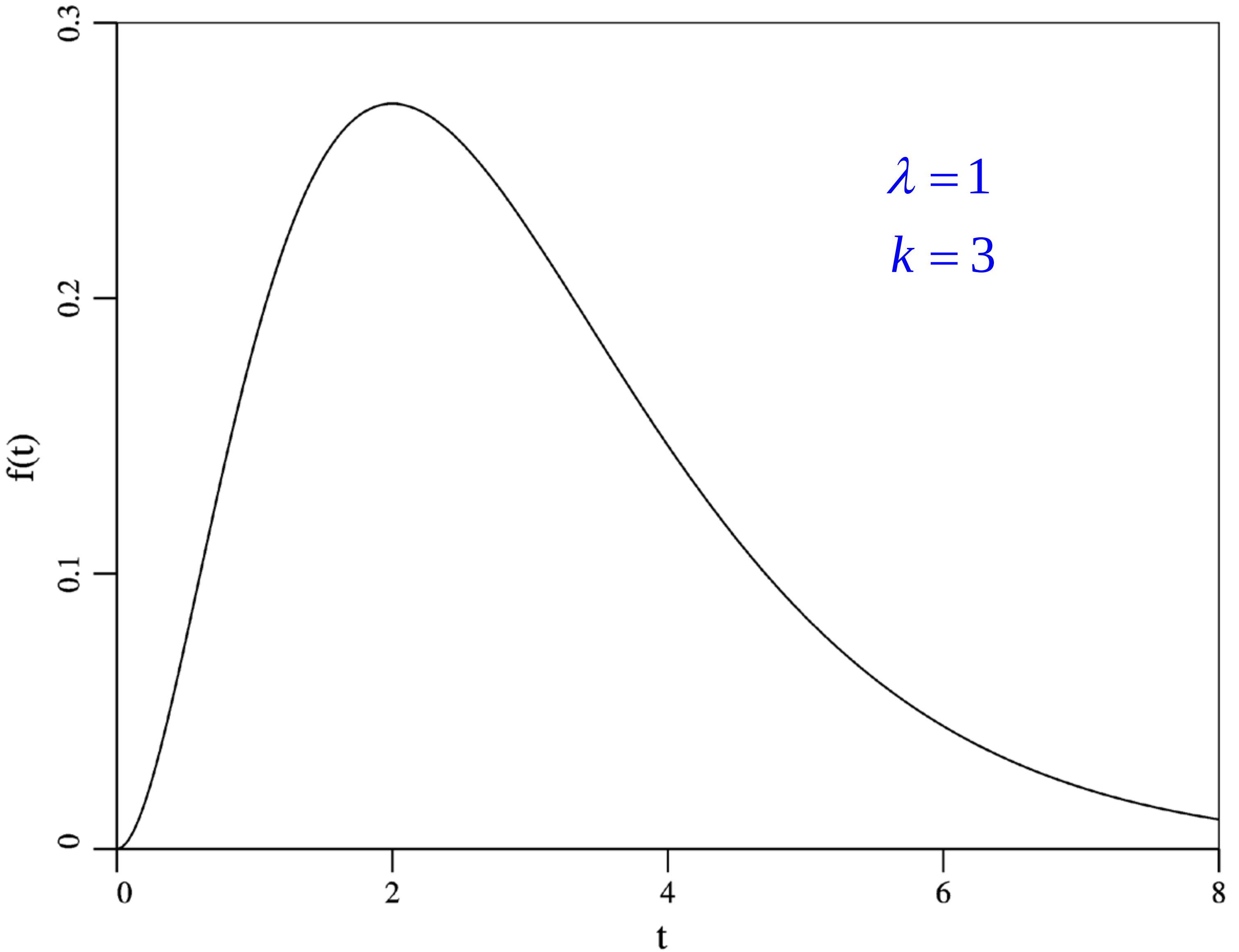
then $F_{T_3}(t) = \mathrm{P}\left[T_3 \le t\right] = 1 - e^{-\lambda t} \sum_{j=0}^{2} \frac{(\lambda t)^j}{j!}$

In general

$$F_{T_k}(t) = \mathrm{P}\left[T_k \le t\right] = 1 - e^{-\lambda t} \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!}$$

is the Gamma distribution

Gamma Distribution



Gamma Distribution: Example

As in the previous example, let us assume that earthquakes in a certain region occur on average once every 50 years and that the number of earthquakes in any time interval follows a Poisson distribution. Under these conditions, what is the probability that less than 150 years will pass before two or more earthquakes occur?

Solution: Let T_2 be the time to the occurrence of the second earthquake. Then, since earthquakes occur according to a Poisson process, T_2 must follow a Gamma distribution with $k = 2$ and $\lambda = 1/50$, which gives us

$$\mathrm{P}\left[T_2 < 150\right] = F_{T_2}\left(150\right) = 1 - e^{-150/50}\left(1 + \frac{\left(150/50\right)^1}{1!}\right) = 0.801$$

Gamma Distribution: Example

Note that since this is a Poisson process, the same result is obtained by noticing that

$$P\left[T_2 < 150\right] = P\left[N_{150} \geq 2\right]$$

so that

$$P\left[N_{150} \geq 2\right] = 1 - P\left[N_{150} < 2\right] = 1 - P\left[N_{150} = 0\right] - P\left[N_{150} = 1\right]$$

$$= 1 - e^{-150/50} - e^{-150/50}\left(\frac{(150/50)^1}{1!}\right)$$

$$= 0.801$$

which is one way of determining the CDF for the Gamma distribution (as we saw earlier).

Weibull Distribution

- generalization of the exponential distribution
- now has two parameters for more flexibility in distribution shape
- common lifetime model
- improving system if $\beta < 1$
- degrading system if $\beta > 1$
- constant failure probability (exponential) if $\beta = 1$

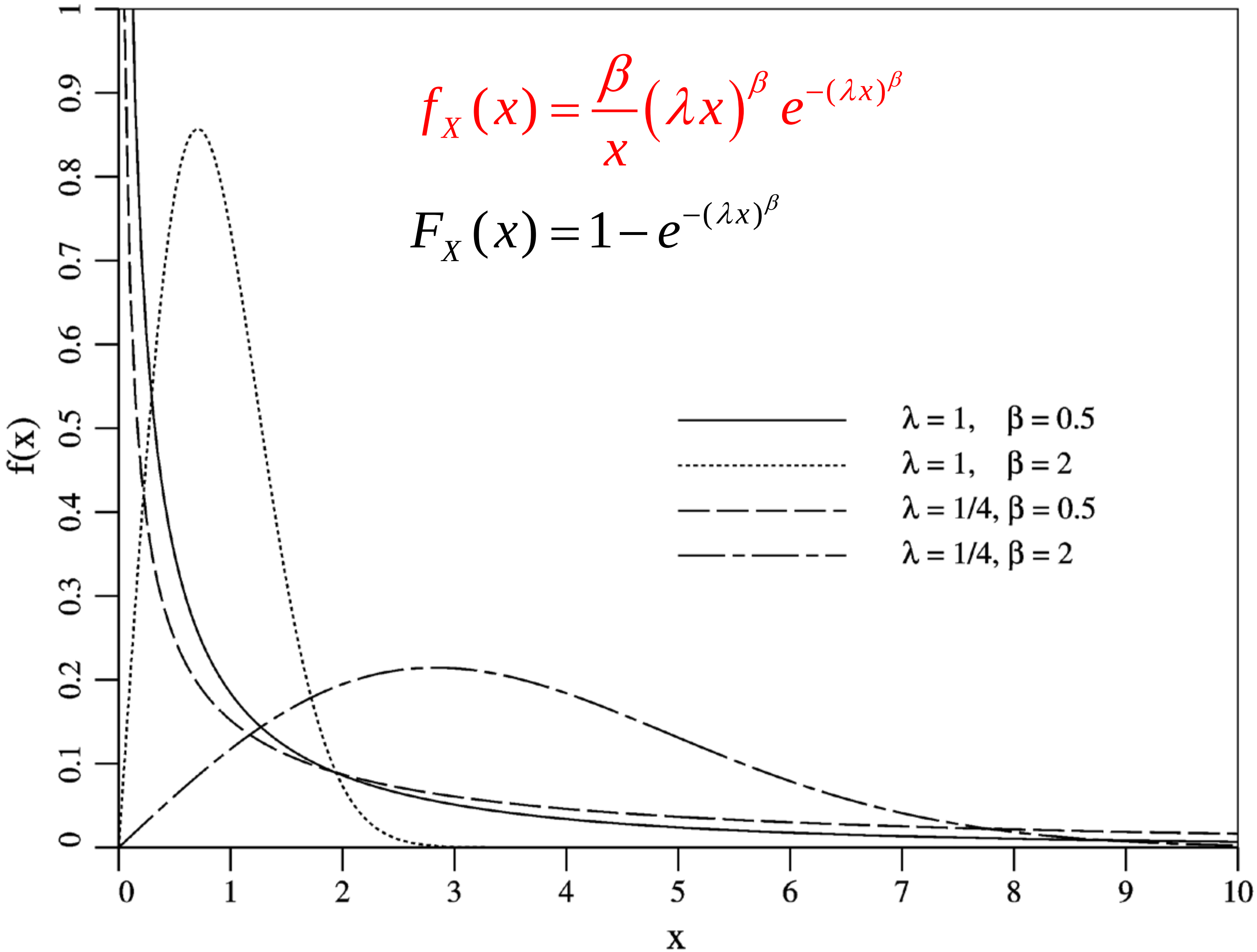
$$f_X(x) = \frac{\beta}{x}\left(\lambda x\right)^{\beta} e^{-(\lambda x)^{\beta}}$$

$$F_X(x) = 1 - e^{-(\lambda x)^{\beta}}$$

Weibull Distribution

$$f_X(x) = \frac{\beta}{x}(\lambda x)^{\beta} e^{-(\lambda x)^{\beta}}$$

$$F_X(x) = 1 - e^{-(\lambda x)^{\beta}}$$



Weibull Distribution: Example

The time to 90% consolidation of a sample of a certain clay has a Weibull distribution with $\beta = 0.5$. A significant number of tests have shown that 81% of clay samples reach 90% consolidation in under 5516 hours. What is the median time to attain 90% consolidation?

Solution: Let X be the time until a clay sample reaches 90% consolidation. Then we are told that X follows a Weibull distribution with $\beta = 0.5$. We first need to compute the other Weibull parameter, λ . To do this we make use of the fact that we know $\mathrm{P}\left[X \le 5516\right] = 0.81$, and since $\mathrm{P}\left[X \le 5516\right] = F_X(5516)$ we have

$$F\left(5516\right) = 1 - e^{-\left(5516\lambda\right)^{0.5}} = 0.81$$

$$e^{-\left(5516\lambda\right)^{0.5}} = 0.19$$

$$\lambda = 1/2000$$

Weibull Distribution: Example

We are now looking for the median, $\tilde{x}$, which is the point which divides the distribution in half. That is, we want to find $\tilde{x}$ such that $F\left(\tilde{x}\right)=0.5$. In other words,

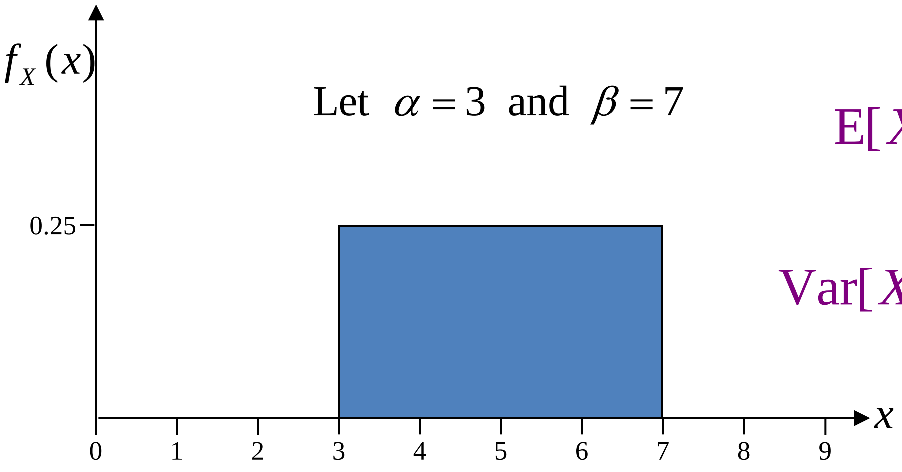
$$1-\exp\left\{-\left(\frac{\tilde{x}}{2000}\right)^{0.5}\right\}=0.5 \quad \Rightarrow \quad \tilde{x}=960.9 \text{ hours}$$

Uniform Distribution

UNIFORM $$f_X(x) = \frac{1}{\beta - \alpha} \qquad \alpha \le x \le \beta$$

The uniform distribution is useful in representing random variables which have known *upper* and *lower* bounds and which have equal likelihood of occurring anywhere between these bounds.

The distribution makes no assumptions regarding preferential likelihood of the random variable since all possible values are equally likely.



Let $\alpha = 3$ and $\beta = 7$

$$\left.\begin{aligned} \mathrm{E}[X] &= \mu_X = \frac{\alpha+\beta}{2} \\ \mathrm{Var}[X] &= \sigma_X^2 = \frac{(\beta-\alpha)^2}{12} \end{aligned}\right\} \text{alternative notation}$$

Normal Distribution

The best known Probability Density Function is the Normal or Gaussian distribution.

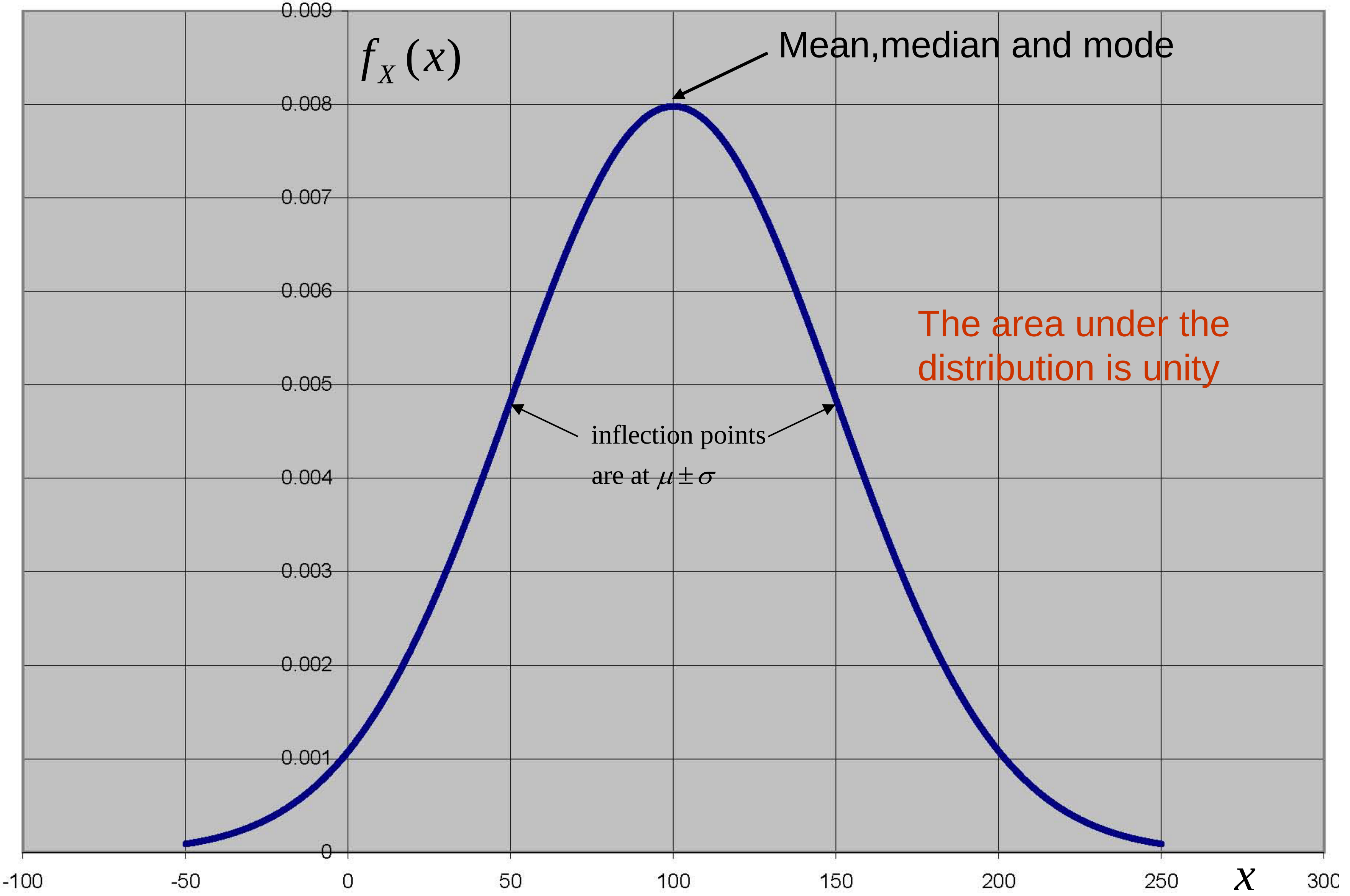
Let X be a normally distributed random variable with mean and standard deviation given by μ_X and σ_X. In this case the PDF is given by:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}$$

$$\mathrm{E}[X] = \mu_X$$
$$\mathrm{Var}[X] = \sigma_X^2$$

NORMAL DISTRIBUTION

$$\mu_X = 100 \qquad \sigma_X = 50$$



NORMAL DISTRIBUTION

To compute probabilities:

$$P[X \le a] = \int_{-\infty}^{a} f_X(x)\,dx = \frac{1}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} dx$$

which has no closed form solution. It must be evaluated numerically → provide tables for probabilities.

Problem: we'd need one table for each possible value of μ_X and σ_X .

Solution:

$$P[X \le a] = P\left[\frac{X-\mu_X}{\sigma_X} \le \frac{a-\mu_X}{\sigma_X}\right] = P\left[Z \le \frac{a-\mu_X}{\sigma_X}\right]$$

$$= \Phi\left(\frac{a-\mu_X}{\sigma_X}\right)$$

STANDARD NORMAL DISTRIBUTION

Where $Z = \dfrac{X - \mu_X}{\sigma_X}$ is the *standard normal*.

$$\mathrm{E}[Z] = \mathrm{E}\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X}\mathrm{E}[X - \mu_X] = 0$$

$$\mathrm{Var}[Z] = \mathrm{Var}\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X^2}\mathrm{Var}[X - \mu_X] = \frac{\mathrm{Var}[X]}{\sigma_X^2} = 1$$

Now we need one table for the standard normal, $\mu_Z = 0$ and $\sigma_Z = 1$ and we *standardize*:

$$\mathrm{P}[X \le a] = \mathrm{P}\left[\frac{X - \mu_X}{\sigma_X} \le \frac{a - \mu_X}{\sigma_X}\right] = \mathrm{P}\left[Z \le \frac{a - \mu_X}{\sigma_X}\right]$$

$$= \Phi\left(\frac{a - \mu_X}{\sigma_X}\right) = \text{ Standard Normal CDF}$$

STANDARD NORMAL DISTRIBUTION

The Standard Normal distribution is so important that it is commonly given its own symbols:

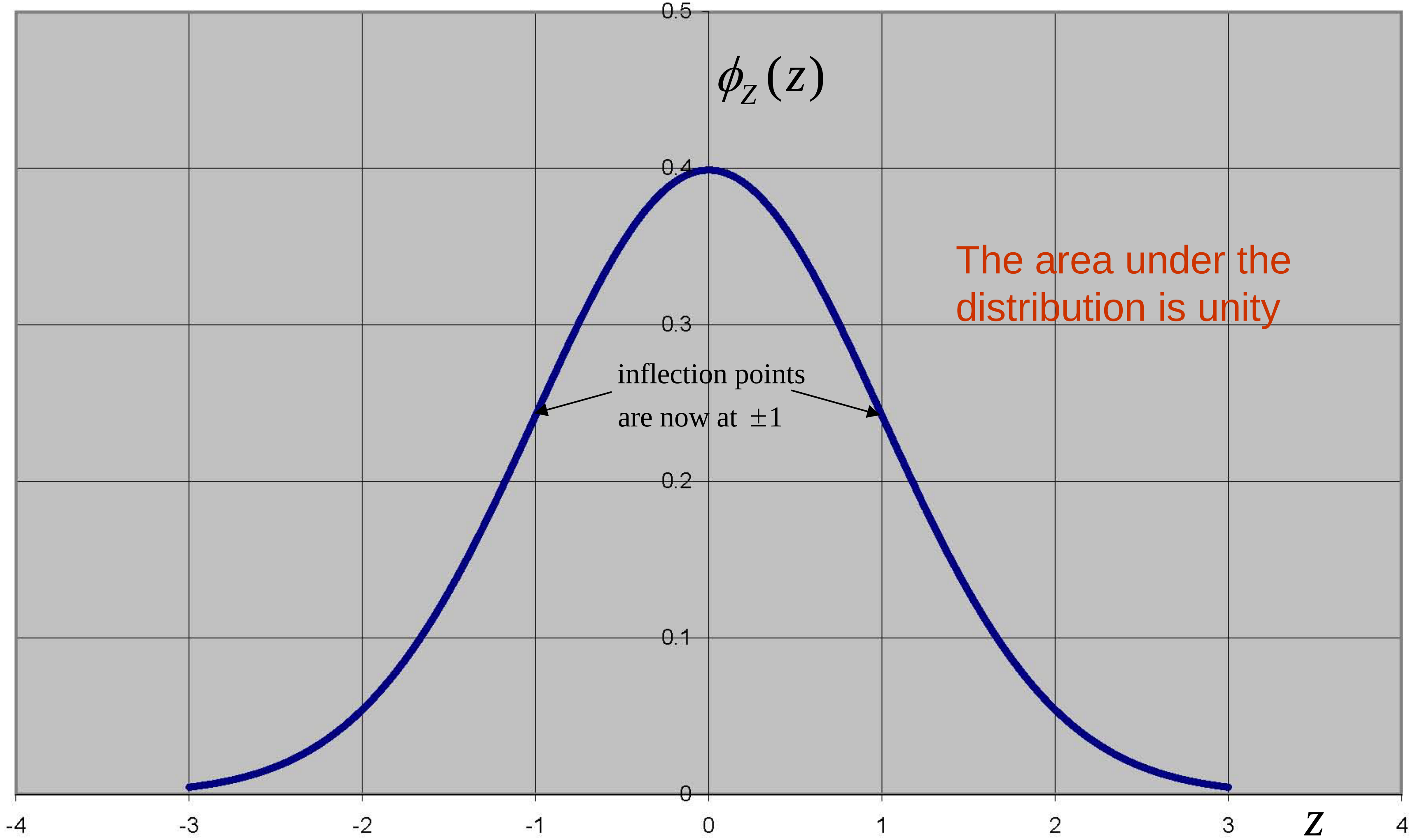
Standardized variable: $Z = \dfrac{X - \mu_X}{\sigma_X}$

Density function: $\phi_Z(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2} z^2\right\}$

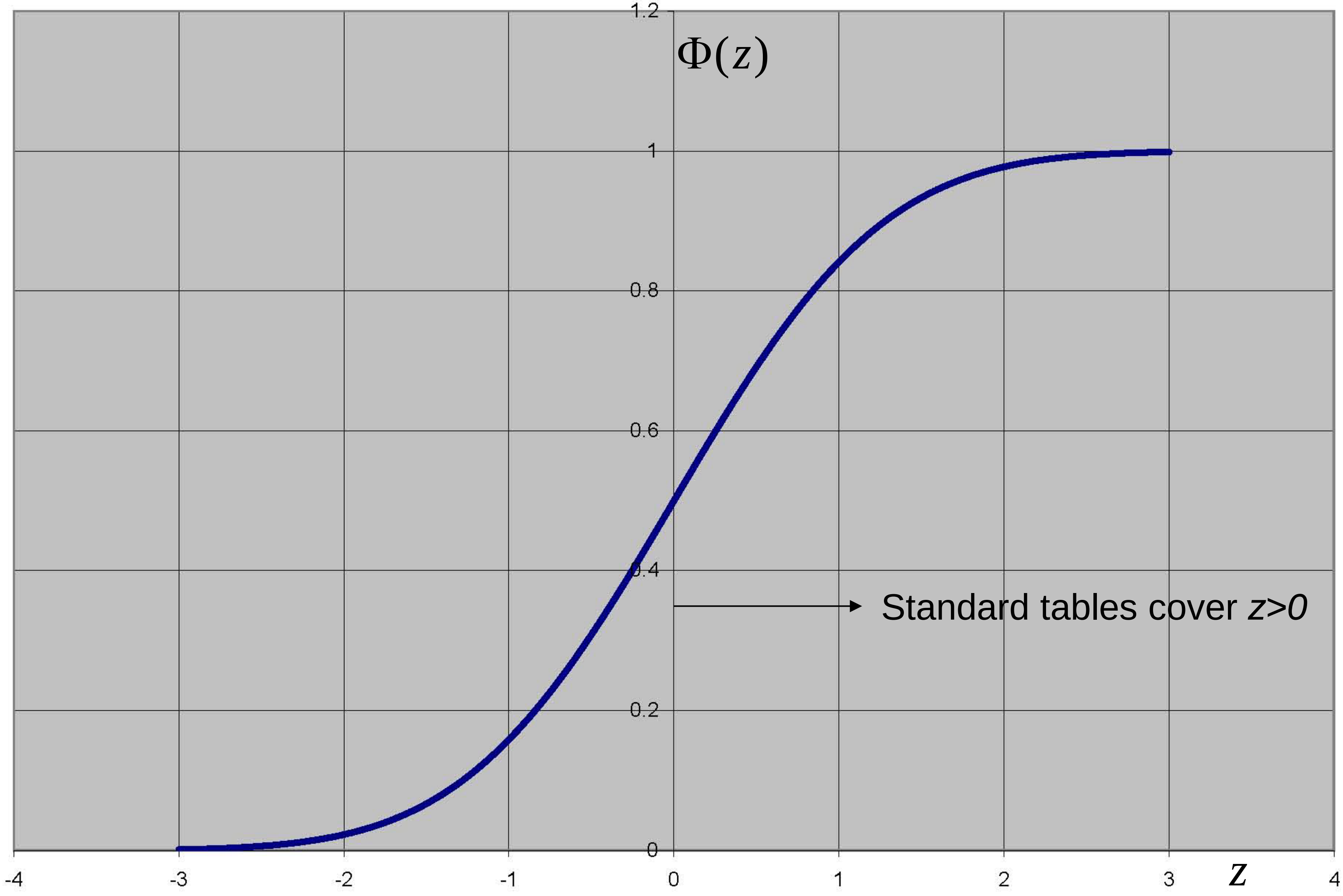
Cumulative Distribution: $\mathrm{P}\left[X \le a\right] = \Phi\left(\dfrac{a - \mu_X}{\sigma_X}\right) = \Phi\left(z\right)$

STANDARD NORMAL DISTRIBUTION

$$\mu_Z = 0 \qquad \sigma_Z = 1$$



Standard Normal Cumulative Distribution Function (CDF)



$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

Standard Normal Function CDF

Table gives $\Phi(z)$ for $z \geq 0$

z	.00	.01	.02	.03	.04	z
0.0	.50000	.50398	.50797	.51196	.51595	0.0
0.1	.53982	.54379	.54775	.55171	.55567	0.1
0.2	.57925	.58316	.58706	.59095	.59483	0.2
0.3	.61791	.62171	.62551	.62930	.63307	0.3
0.4	.65542	.65909	.66275	.66640	.67003	0.4
0.5	.69146	.69497	.69846	.70194	.70540	0.5
0.6	.72574	.72906	.73237	.73565	.73891	0.6
0.7	.75803	.76114	.76423	.76730	.77035	0.7
0.8	.78814	.79102	.79389	.79673	.79954	0.8
0.9	.81593	.81858	.82121	.82381	.82639	0.9
1.0	.84134	.84375	.84613	.84849	.85083	1.0
1.1	.86433	.86650	.86864	.87076	.87285	1.1
1.2	.88493	.88686	.88876	.89065	.89251	1.2
1.3	.90319	.90490	.90658	.90824	.90987	1.3
1.4	.91924	.92073	.92219	.92364	.92506	1.4
1.5	.93319	.93447	.93574	.93699	.93821	1.5
1.6	.94520	.94630	.94738	.94844	.94949	1.6
1.7	.95543	.95636	.95728	.95818	.95907	1.7
1.8	.96406	.96485	.96562	.96637	.96711	1.8
1.9	.97128	.97193	.97257	.97319	.97381	1.9
2.0	.97724	.97778	.97830	.97882	.97932	2.0
2.1	.98213	.98257	.98299	.98341	.98382	2.1
2.2	.98609	.98644	.98679	.98712	.98745	2.2
2.3	.98927	.98955	.98982	$.9^2 0096$	$.9^2 0358$	2.3
2.4	$.9^2 1802$	$.9^2 2023$	$.9^2 2239$	$.9^2 2450$	$.9^2 2656$	2.4
2.5	$.9^2 3790$	$.9^2 3963$	$.9^2 4132$	$.9^2 4296$	$.9^2 4457$	2.5
2.6	$.9^2 5338$	$.9^2 5472$	$.9^2 5603$	$.9^2 5730$	$.9^2 5854$	2.6
2.7	$.9^2 6533$	$.9^2 6635$	$.9^2 6735$	$.9^2 6833$	$.9^2 6928$	2.7
2.8	$.9^2 7444$	$.9^2 7522$	$.9^2 7598$	$.9^2 7672$	$.9^2 7744$	2.8
2.9	$.9^2 8134$	$.9^2 8192$	$.9^2 8249$	$.9^2 8305$	$.9^2 8358$	2.9
3.0	$.9^2 8650$	$.9^2 8693$	$.9^2 8736$	$.9^2 8777$	$.9^2 8817$	3.0
3.1	$.9^3 0323$	$.9^3 0645$	$.9^3 0957$	$.9^3 1259$	$.9^3 1552$	3.1
3.2	$.9^3 3128$	$.9^3 3363$	$.9^3 3590$	$.9^3 3810$	$.9^3 4023$	3.2
3.3	$.9^3 5165$	$.9^3 5335$	$.9^3 5499$	$.9^3 5657$	$.9^3 5811$	3.3
3.4	$.9^3 6630$	$.9^3 6751$	$.9^3 6868$	$.9^3 6982$	$.9^3 7091$	3.4
3.5	$.9^3 7673$	$.9^3 7759$	$.9^3 7842$	$.9^3 7922$	$.9^3 7999$	3.5
3.6	$.9^3 8408$	$.9^3 8469$	$.9^3 8526$	$.9^3 8582$	$.9^3 8636$	3.6
3.7	$.9^3 8922$	$.9^3 8963$	$.9^4 0038$	$.9^4 0426$	$.9^4 0799$	3.7
3.8	$.9^4 2765$	$.9^4 3051$	$.9^4 3327$	$.9^4 3592$	$.9^4 3848$	3.8
3.9	$.9^4 5190$	$.9^4 5385$	$.9^4 5572$	$.9^4 5752$	$.9^4 5925$	3.9
4.0	$.9^4 6832$	$.9^4 6964$	$.9^4 7090$	$.9^4 7211$	$.9^4 7327$	4.0

Notes:

1) for $z = i.jk$, where i, j, and k are digits, enter table at line $i.j$ under column $.0k$ (next page for $k \geq 5$).
2) $0.9^4 7327$ is short for 0.99997327, etc.

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

Standard Normal Function CDF

For z < 0

$$\Phi(z) = 1 - \Phi(-z)$$

z	.05	.06	.07	.08	.09	z
0.0	.51993	.52392	.52790	.53188	.53585	0.0
0.1	.55961	.56355	.56749	.57142	.57534	0.1
0.2	.59870	.60256	.60641	.61026	.61409	0.2
0.3	.63683	.64057	.64430	.64802	.65173	0.3
0.4	.67364	.67724	.68082	.68438	.68793	0.4
0.5	.70884	.71226	.71566	.71904	.72240	0.5
0.6	.74215	.74537	.74857	.75174	.75490	0.6
0.7	.77337	.77637	.77935	.78230	.78523	0.7
0.8	.80233	.80510	.80784	.81057	.81326	0.8
0.9	.82894	.83147	.83397	.83645	.83891	0.9
1.0	.85314	.85542	.85769	.85992	.86214	1.0
1.1	.87492	.87697	.87899	.88099	.88297	1.1
1.2	.89435	.89616	.89795	.89972	.90147	1.2
1.3	.91149	.91308	.91465	.91620	.91773	1.3
1.4	.92647	.92785	.92921	.93056	.93188	1.4
1.5	.93942	.94062	.94179	.94294	.94408	1.5
1.6	.95052	.95154	.95254	.95352	.95448	1.6
1.7	.95994	.96079	.96163	.96246	.96327	1.7
1.8	.96784	.96855	.96925	.96994	.97062	1.8
1.9	.97441	.97500	.97558	.97614	.97670	1.9
2.0	.97981	.98030	.98077	.98123	.98169	2.0
2.1	.98422	.98461	.98499	.98537	.98573	2.1
2.2	.98777	.98808	.98839	.98869	.98898	2.2
2.3	$.9^2 0613$	$.9^2 0862$	$.9^2 1105$	$.9^2 1343$	$.9^2 1575$	2.3
2.4	$.9^2 2857$	$.9^2 3053$	$.9^2 3244$	$.9^2 3430$	$.9^2 3612$	2.4
2.5	$.9^2 4613$	$.9^2 4766$	$.9^2 4915$	$.9^2 5059$	$.9^2 5201$	2.5
2.6	$.9^2 5975$	$.9^2 6092$	$.9^2 6207$	$.9^2 6318$	$.9^2 6427$	2.6
2.7	$.9^2 7020$	$.9^2 7109$	$.9^2 7197$	$.9^2 7282$	$.9^2 7364$	2.7
2.8	$.9^2 7814$	$.9^2 7881$	$.9^2 7947$	$.9^2 8011$	$.9^2 8073$	2.8
2.9	$.9^2 8411$	$.9^2 8461$	$.9^2 8511$	$.9^2 8558$	$.9^2 8605$	2.9
3.0	$.9^2 8855$	$.9^2 8893$	$.9^2 8929$	$.9^2 8964$	$.9^2 8999$	3.0
3.1	$.9^3 1836$	$.9^3 2111$	$.9^3 2378$	$.9^3 2636$	$.9^3 2886$	3.1
3.2	$.9^3 4229$	$.9^3 4429$	$.9^3 4622$	$.9^3 4809$	$.9^3 4990$	3.2
3.3	$.9^3 5959$	$.9^3 6102$	$.9^3 6241$	$.9^3 6375$	$.9^3 6505$	3.3
3.4	$.9^3 7197$	$.9^3 7299$	$.9^3 7397$	$.9^3 7492$	$.9^3 7584$	3.4
3.5	$.9^3 8073$	$.9^3 8145$	$.9^3 8215$	$.9^3 8282$	$.9^3 8346$	3.5
3.6	$.9^3 8688$	$.9^3 8738$	$.9^3 8787$	$.9^3 8833$	$.9^3 8878$	3.6
3.7	$.9^4 1158$	$.9^4 1504$	$.9^4 1837$	$.9^4 2158$	$.9^4 2467$	3.7
3.8	$.9^4 4094$	$.9^4 4330$	$.9^4 4558$	$.9^4 4777$	$.9^4 4987$	3.8
3.9	$.9^4 6092$	$.9^4 6252$	$.9^4 6406$	$.9^4 6554$	$.9^4 6696$	3.9
4.0	$.9^4 7439$	$.9^4 7546$	$.9^4 7649$	$.9^4 7748$	$.9^4 7843$	4.0

Notes:

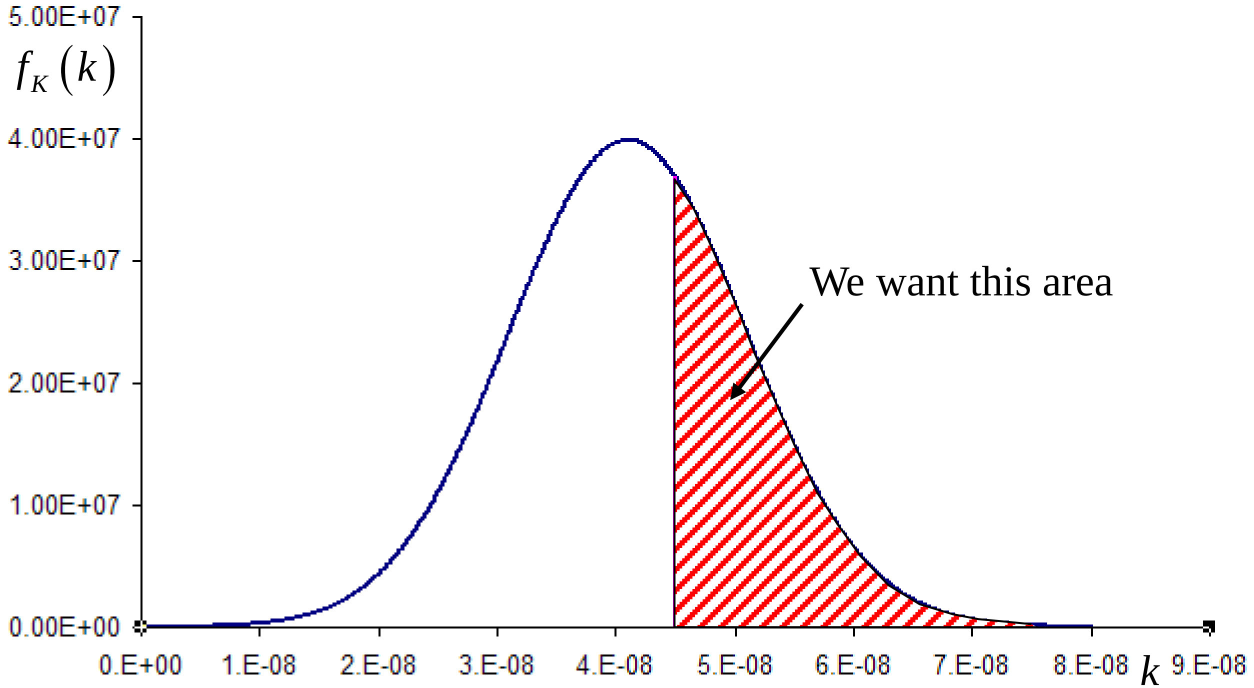
1) for $z = i.jk$, where i, j, and k are digits, enter table at line $i.j$ under column $.0k$ (next page for $k \geq 5$).
2) $0.9^4 7439$ is short for 0.99997439, etc.
3) $\Phi(-z) = 1 - \Phi(z)$

Example calculations using the Standard Normal Function CDF

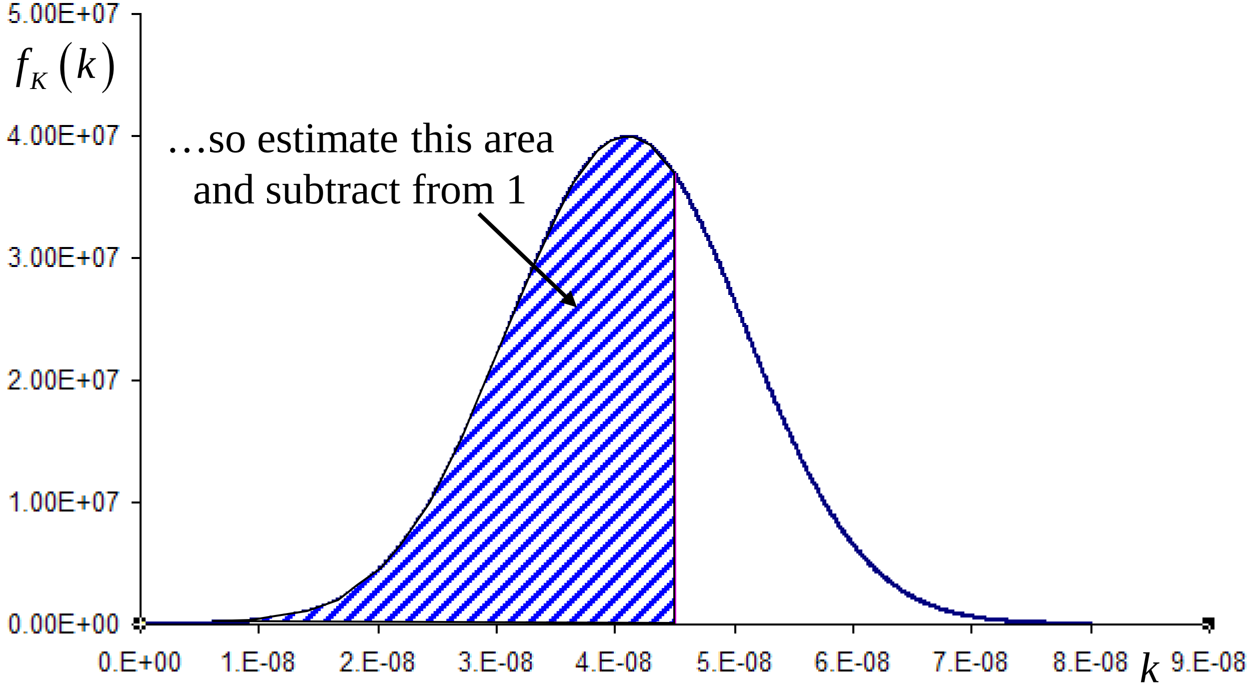
Example 1:

Permeability measurements have indicated that *k* is normally distributed with the properties: $\mu_k = 4.1\times10^{-8}\,\text{m/s}$ and $\sigma_k = 1\times10^{-8}\,\text{m/s}$

What is the probability that $k > 4.5\times10^{-8}\,\text{m/s}$?



…but tables only give area to the <u>left</u> of a given point…



$$
\begin{aligned}
P[k > 4.5\times10^{-8}] &= 1 - P[k \le 4.5\times10^{-8}] \\
&= 1 - \Phi\left(\frac{4.5\times10^{-8} - 4.1\times10^{-8}}{1\times10^{-8}}\right) \\
&= 1 - \Phi(0.4) \\
&= 1 - 0.65542 \\
&= 0.3446 \quad (34\%)
\end{aligned}
$$

The Reliability Index

The Reliability Index is a measure of the margin of safety in "standard deviation units".

For example, if dealing with a normally distributed Factor of Safety (where *FS=1* implies failure), the reliability Index is given by:

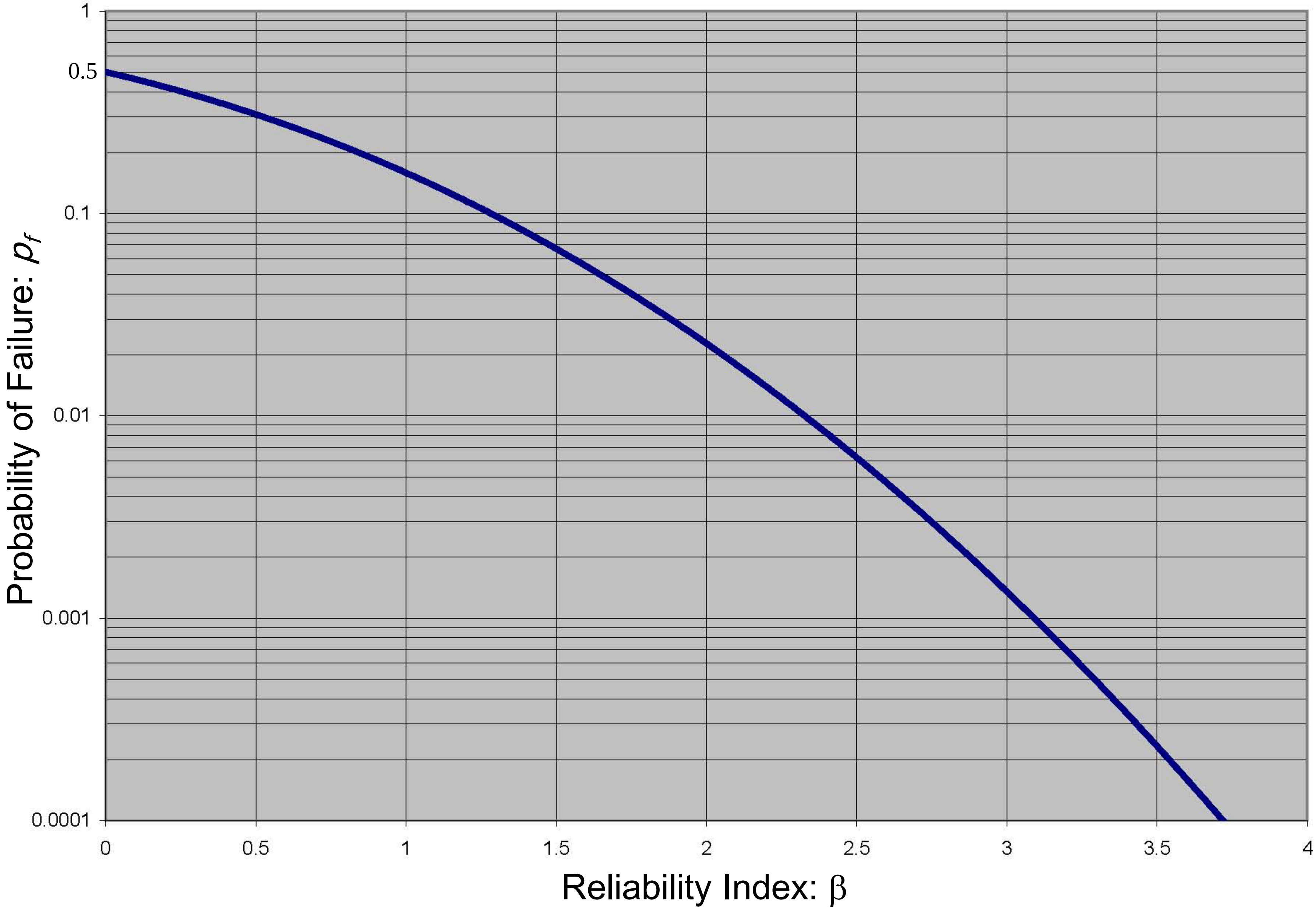
$$\beta = \frac{\mu_{FS} - 1}{\sigma_{FS}}$$

If the Factor of Safety is lognornal, the reliability Index is given by:

$$\beta = \frac{\mu_{\ln(FS)}}{\sigma_{\ln(FS)}}$$

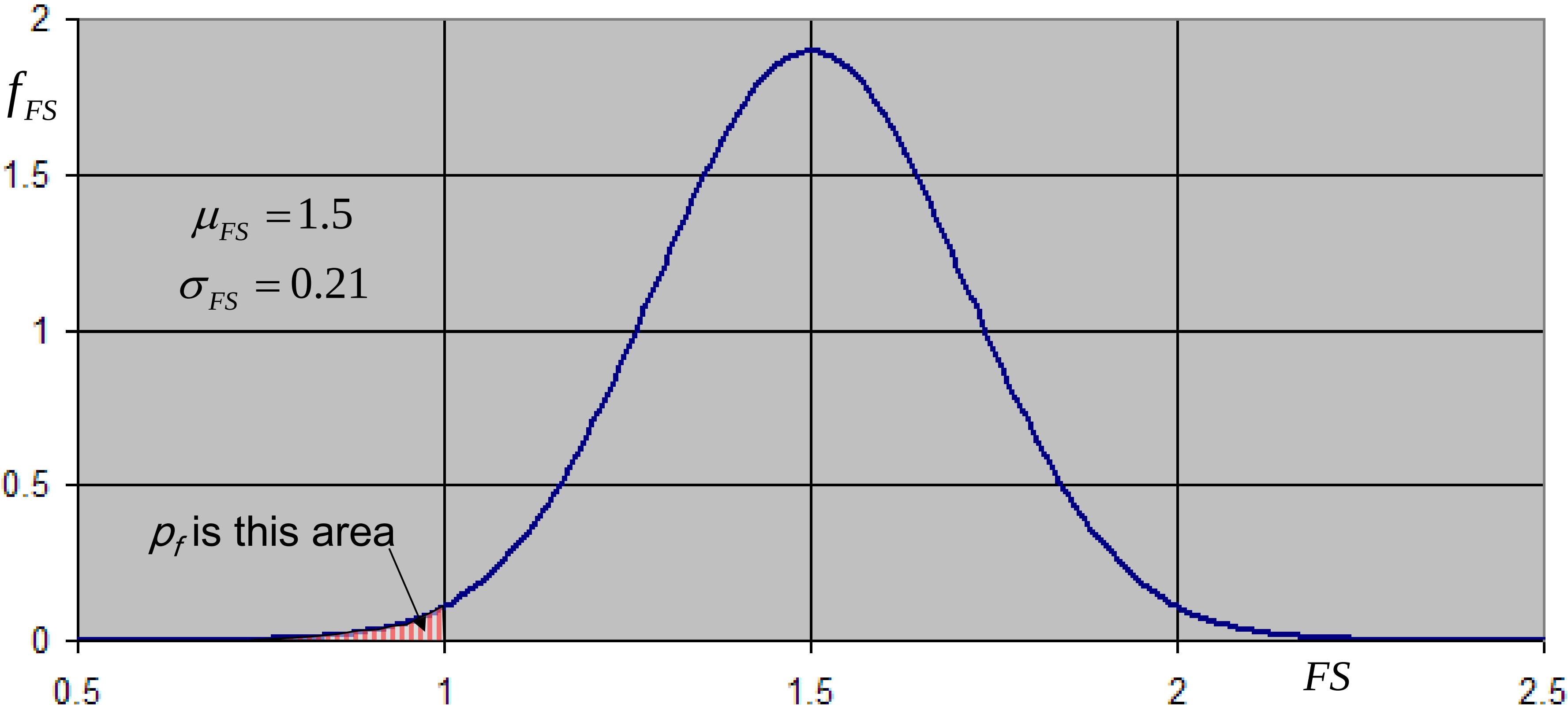
For normally distributed random variables, the "reliability index" (β) is uniquely related to the "probability of failure" (p_f) through the expression:

$$p_f = 1 - \Phi(\beta)$$



Probability of Failure vs. Reliability Index for a Normal Distribution

Consider a normal distribution of the Factor of Safety (FS)



β is this distance ÷ the standard deviation

$$\beta = \frac{(1.5-1)}{0.21} = 2.38 \qquad p_f = 1 - \Phi(2.38) = 1 - 0.991343 = 0.0087$$

Lognormal Distribution

Let X be a lognormally distributed random variable with a mean and standard deviation given by μ_X and σ_X.

Let the Coefficient of Variation $v_X = \dfrac{\sigma_X}{\mu_X}$

If X is lognormally distributed, this means that $\ln X$ is normally distributed. Let the mean and standard deviation of this underlying normal distribution of $\ln X$ be given by $\mu_{\ln X}$ and $\sigma_{\ln X}$ respectively.

The PDF for a lognormal distribution is given by:

$$f_X(x) = \frac{1}{x\sigma_{\ln X}\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu_{\ln X}}{\sigma_{\ln X}}\right)^2\right\}$$

If the mean and standard deviation of the lognormal random variable x are μ_X and σ_X, then the mean and standard deviation of the underlying normal distribution of $\ln x$ are given by:

$$\mu_{\ln X} = \ln \mu_X - \frac{1}{2}\ln\left\{1 + v_X{}^2\right\}$$

$$\sigma_{\ln X} = \sqrt{\ln\left\{1 + v_X{}^2\right\}}$$

Lognormal Distribution

To compute probabilities:

1. convert to normal by taking logarithms on both sides of the inequality,
2. standardize by subtracting mean and dividing by standard deviation on both sides of inequality,
3. look up probability in standard normal tables

$$\mathrm{P}\left[X \leq a\right] = \mathrm{P}\left[\ln X \leq \ln a\right] = \mathrm{P}\left[\frac{\ln X - \mu_{\ln X}}{\sigma_{\ln X}} \leq \frac{\ln a - \mu_{\ln X}}{\sigma_{\ln X}}\right]$$

$$= \mathrm{P}\left[Z \leq \frac{\ln a - \mu_{\ln X}}{\sigma_{\ln X}}\right] = \Phi\left(\frac{\ln a - \mu_{\ln X}}{\sigma_{\ln X}}\right)$$

Going in the other direction....

$$\mu_X = \exp\left(\mu_{\ln X} + \frac{1}{2}\sigma^2_{\ln X}\right)$$

$$\sigma_X = \mu_X \sqrt{\exp\left(\sigma^2_{\ln X}\right) - 1}$$

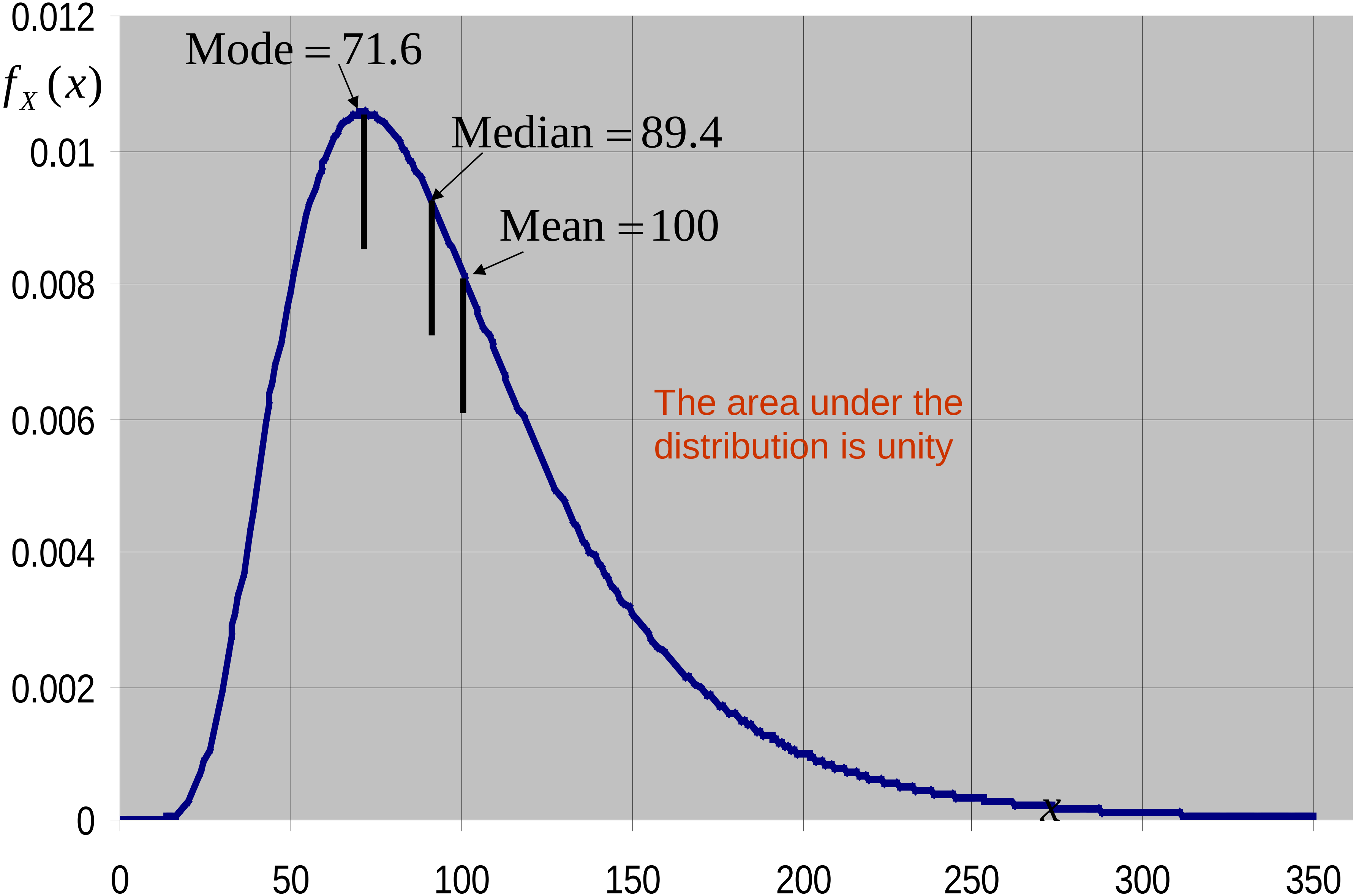
Further relationships involving the **lognormal distribution**:

$$\text{Median}_X = \exp\left(\mu_{\ln X}\right)$$

$$\text{Mode}_X = \exp\left(\mu_{\ln X} - \sigma^2_{\ln X}\right)$$

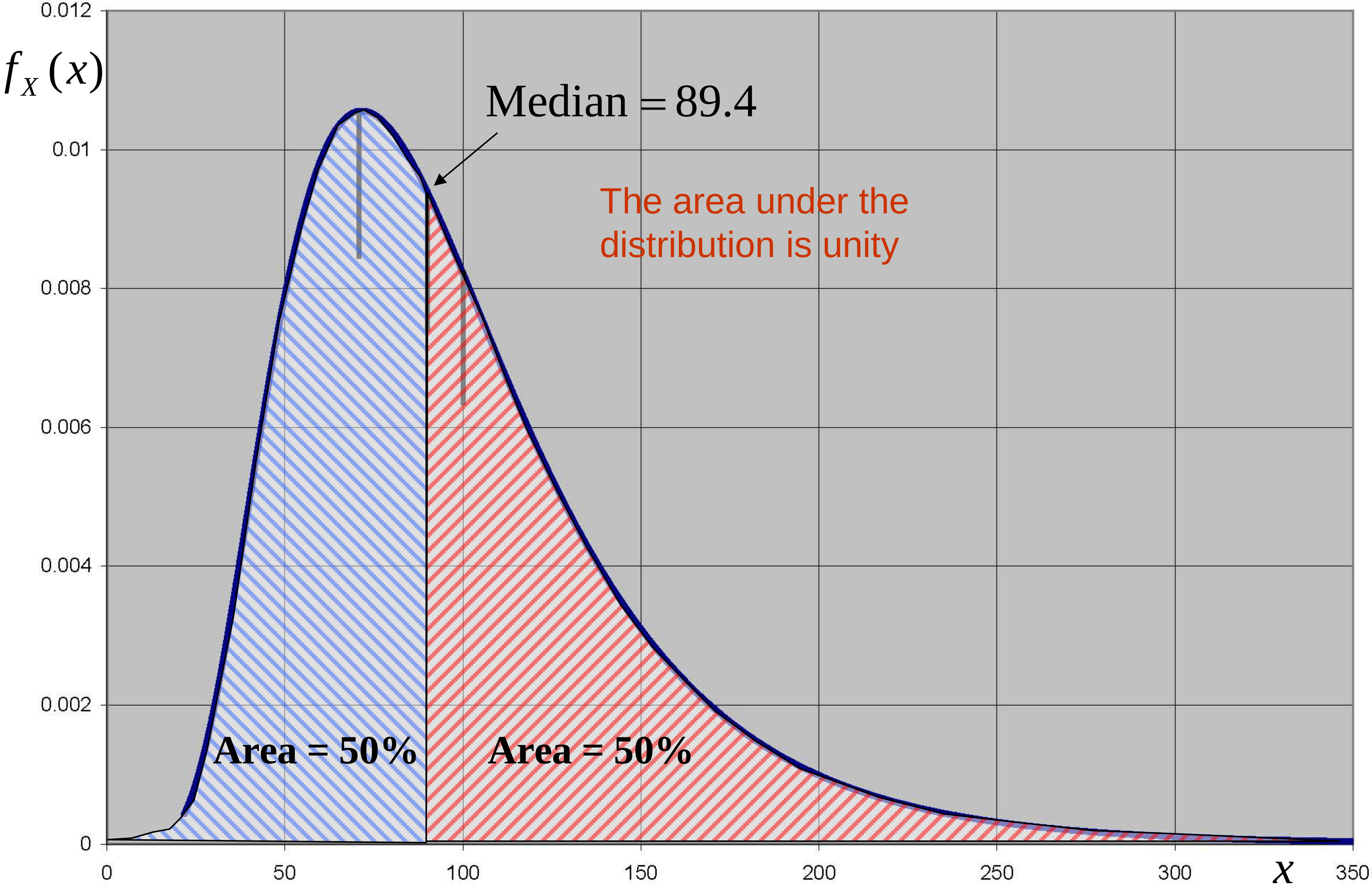
LOGNORMAL DISTRIBUTION

$$\mu_X = 100 \quad \sigma_X = 50$$



SIGNIFICANCE OF THE MEDIAN IN A LOGNORMAL DISTRIBUTION

$$\mu_X = 100 \qquad \sigma_X = 50$$



Example 2:

Permeability measurements have indicated that *k* is lognormally distributed with the properties: $\mu_k = 4.1\times 10^{-8}\,\mathrm{m/s}$ and $\sigma_k = 1\times 10^{-8}\,\mathrm{m/s}$

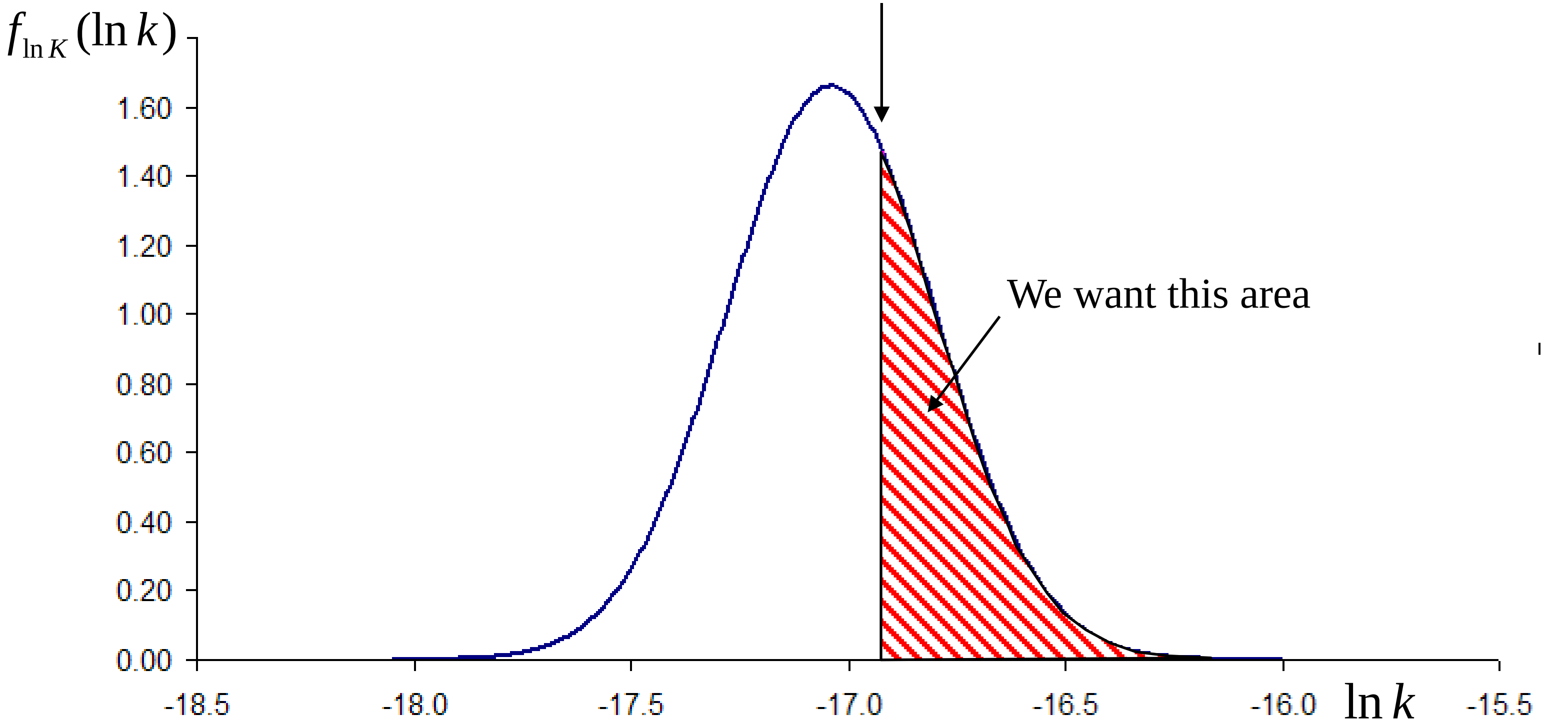
What is the probability that $k > 4.5\times 10^{-8}\,\mathrm{m/s}$?

First find the properties of the underlying <u>normal</u> distribution of $\ln k$

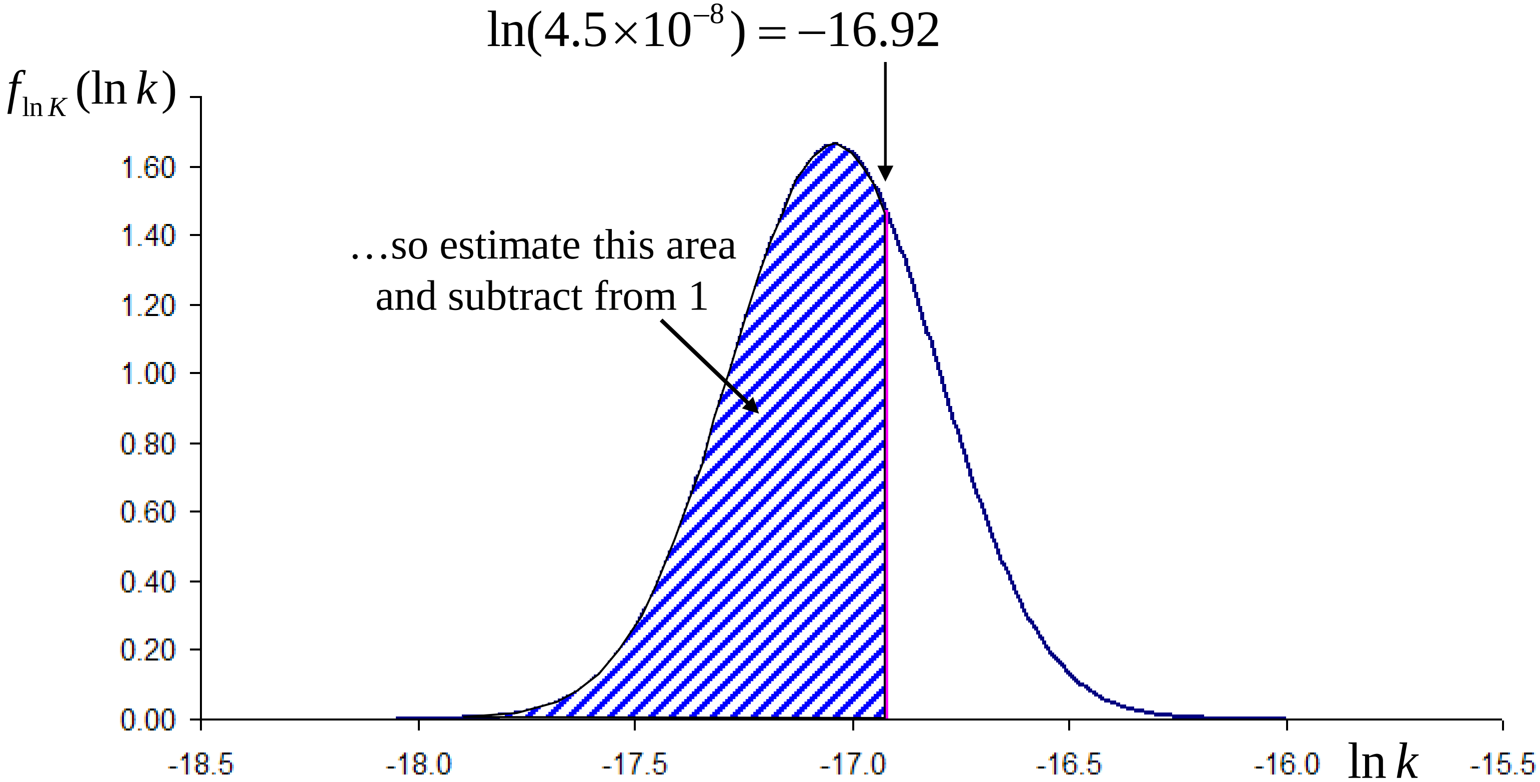
$$\sigma_{\ln k} = \sqrt{\ln\left\{1+\left(\frac{\sigma_k}{\mu_k}\right)^2\right\}} = \sqrt{\ln\left\{1+\left(\frac{1}{4.1}\right)^2\right\}} = 0.2404$$

$$\mu_{\ln k} = \ln \mu_k - \frac{1}{2}\sigma_{\ln k}^2 = \ln(4.1\times 10^{-8}) - \frac{1}{2}(0.2404)^2 = -17.0386$$

$\ln(4.5\times10^{-8}) = -16.92$



…but tables only give area to the <u>left</u> of a given point…



$$P[k > 4.5\times10^{-8}] = 1 - P[k \le 4.5\times10^{-8}]$$

$$= 1 - \Phi\left(\frac{-16.92 - (-17.0386)}{0.2404}\right)$$

$$= 1 - \Phi(0.5075)$$

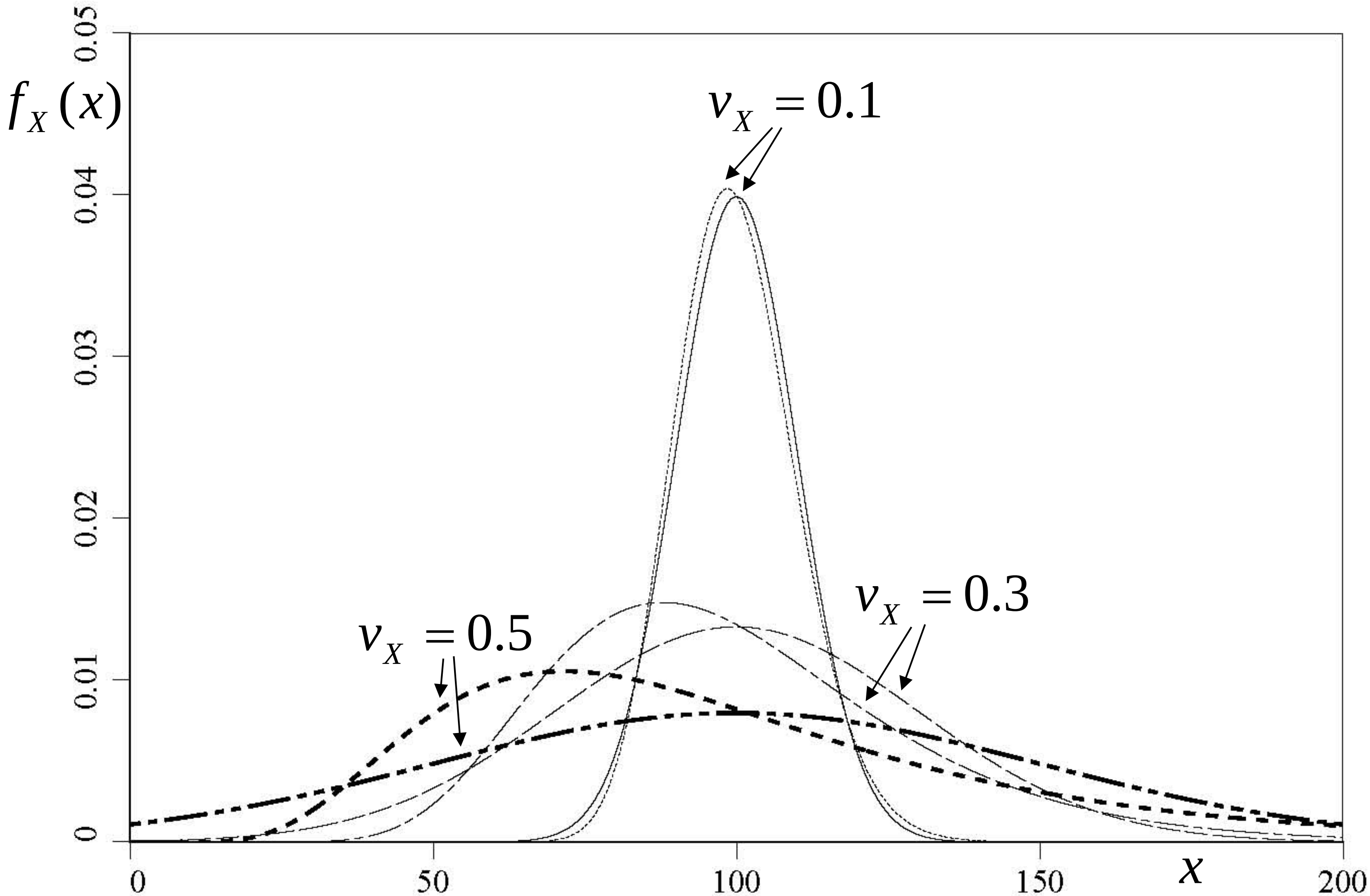
$$\approx 1 - \Phi(0.51)$$

$$= 1 - 0.69497$$

$$= 0.3050 \ (31\%) \quad \text{(was 0.3446 using the normal distribution)}$$

COMPARISON OF NORMAL AND LOGNORMAL DISTRIBUTIONS

$$\mu_X = 100 \text{ (Mean)} \qquad v_X = \frac{\sigma_X}{\mu_X} \text{ (Coefficient of Variation)}$$

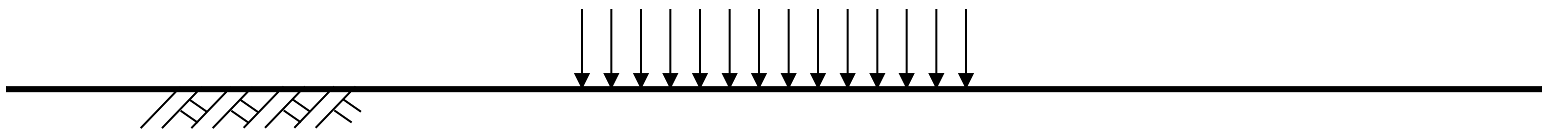


....not much difference for $v_X < 0.3$

Calculations involving a single random variable (SRV)

Example of Bearing Capacity

q



$\phi_u = 0, \;\; c_u \;\; \text{kN/m}^2$

What is the relationship between the Factor of Safety (*FS*) of a conventional bearing capacity calculation (based on the mean strength) and the probability of failure (p_f)?

First perform a **deterministic** calculation

Conventional bearing capacity calculations typically involve high factors of safety of at least 3.

The bearing capacity of an undrained clay is given by the Prandtl equation:

$$\begin{aligned} q_u &= (2+\pi)c_u \\ &= 5.14c_u \end{aligned}$$

where c_u is a design “mean” value of the undrained shear strength.

If $c_u = 100\text{kN/m}^2$, and $FS = 3$ this implies an allowable bearing pressure of:

$$q_{all} = \frac{5.14 \times 100}{3} = 171\text{kN/m}^2$$

Now perform a **probabilistic** calculation

If additional data comes in to indicate that the undrained strength is a lognormal random variable with $\mu_{c_u} = 100\text{kN/m}^2$ and $\sigma_{c_u} = 50\text{kN/m}^2$ what is the probability of the actual bearing capacity being less than the factored deterministic value $P[q_u < 171]$?

q_u is a linear function of the random variable c_u from $q_u = 5.14c_u$

hence

$$\mathrm{E}[q_u] = 5.14\mathrm{E}[c_u] \qquad \text{thus } \mu_{q_u} = 5.14\mu_{c_u} = 514$$

and

$$\mathrm{Var}[q_u] = 5.14^2\,\mathrm{Var}[c_u] \qquad \text{thus } \sigma_{q_u} = 5.14\sigma_{c_u} = 257$$

(Note that $v_{q_u} = v_{c_u} = \frac{1}{2}$)

First find the properties of the underlying <u>normal</u> distribution of $\ln q_u$

$$\sigma_{\ln q_u} = \sqrt{\ln\left\{1+v_{q_u}{}^2\right\}} = \sqrt{\ln\left\{1+\left(\frac{1}{2}\right)^2\right\}} = 0.47$$

$$\mu_{\ln q_u} = \ln \mu_{q_u} - \frac{1}{2}\sigma^2_{\ln q_u} = \ln(514) - \frac{1}{2}(0.47)^2 = 6.13$$

$$P[q_u < 171] = \Phi\left(\frac{\ln 171 - 6.13}{0.47}\right)$$

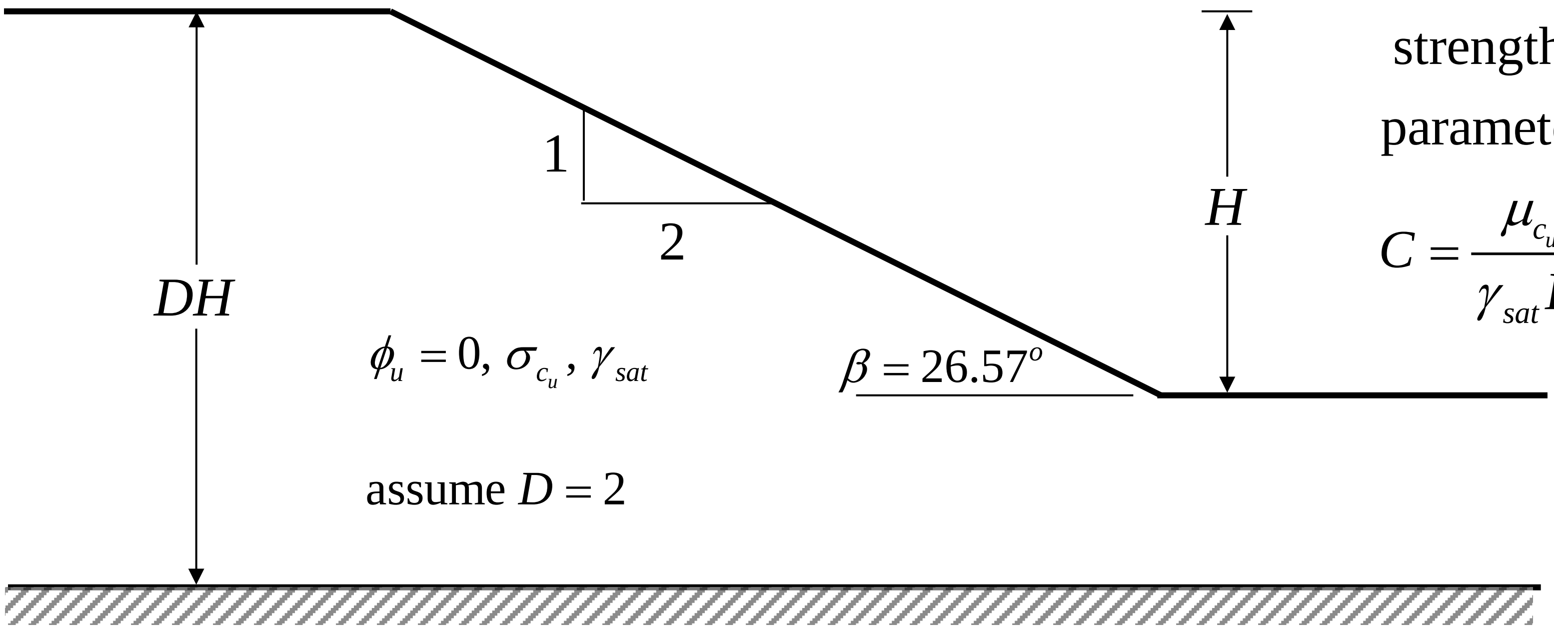
$$= \Phi\left(-2.10\right)$$

$$= 1 - \Phi\left(2.10\right)$$

$$= 1 - 0.982$$

$$= 0.018\ (1.8\%)$$

Example of Slope Stability



Dimensionless strength parameter

$$C = \frac{\mu_{c_u}}{\gamma_{sat} H}$$

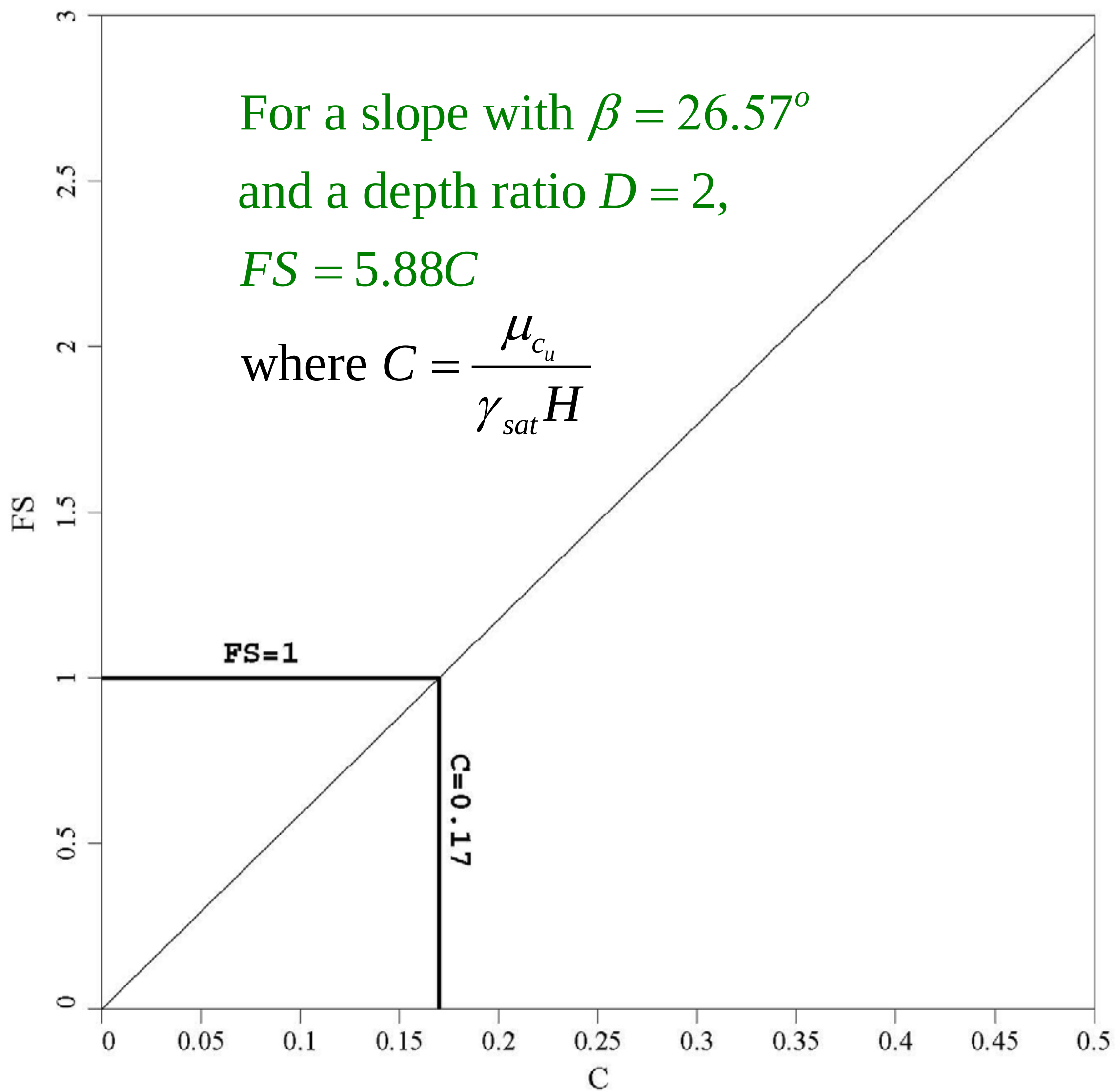
If c_u is treated as a random variable

with mean μ_{c_u} and standard deviation σ_{c_u}

what is the relationship between

FS (based on the mean) and the probability of failure p_f ?

Deterministic calculation



Probabilistic calculation

1) Note that $FS \propto C$ or $FS = KC$,

hence $\mu_{FS} = K\mu_C$ and $v_{FS} = v_C$

2) If C is lognormal, then so is FS.

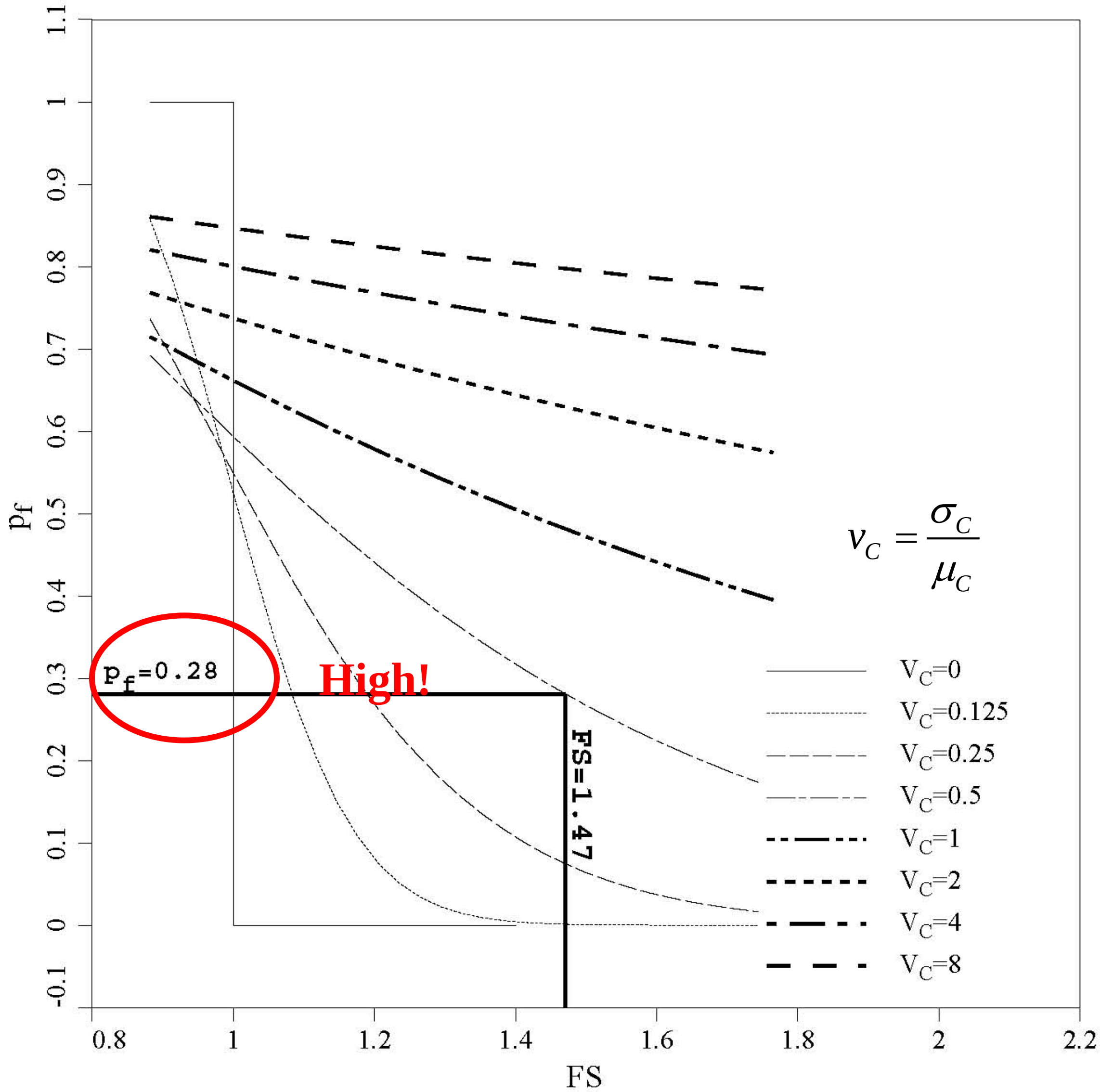
Compute the underlying normal properties $\sigma_{\ln FS}$ and $\mu_{\ln FS}$

3) Find the probability $P\left[FS < 1\right]$ from standard tables.

If FS is lognormal, $P\left[FS < 1\right] = P\left[\ln FS < \ln 1\right]$

$$= \Phi\left[\frac{\ln 1 - \mu_{\ln FS}}{\sigma_{\ln FS}}\right] = \Phi\left[-\frac{\mu_{\ln FS}}{\sigma_{\ln FS}}\right]$$

If FS is normal, $P\left[FS < 1\right] = \Phi\left[\frac{1 - \mu_{FS}}{\sigma_{FS}}\right]$



The MEAN strength may be too optiminstic….

Extreme Value Distributions

Most systems fail due to a combination of extremes in load and/or resistance, so extreme value distributions are very useful in assessing system reliability.

Let $Y_n = \max\left(X_1, X_2, \ldots, X_n\right)$

and $Y_1 = \min\left(X_1, X_2, \ldots, X_n\right)$

We are interested in determining the distributions of Y_n and Y_1

Exact Extreme Value Distributions: Maximum

Since Y_n is the maximum of the X's, then the event $Y_n \leq y$ implies that all of the X's are also less than y. In other words,

$$\mathrm{P}\left[Y_n \leq y\right] = \mathrm{P}\left[X_1 \leq y \cap X_2 \leq y \cap \cdots \cap X_n \leq y\right]$$

If it is assumed that the X's are independent and identically distributed, then

$$\begin{aligned} F_{Y_n}(y) &= \mathrm{P}\left[Y_n \leq y\right] = \mathrm{P}\left[X_1 \leq y\right]\mathrm{P}\left[X_2 \leq y\right]\cdots\mathrm{P}\left[X_n \leq y\right] \\ &= \left[F_X(y)\right]^n \end{aligned}$$

$$f_{Y_n}(y) = \frac{dF_{Y_n}(y)}{dy} = n\, f_X(y)\left[F_X(y)\right]^{n-1}$$

Exact Extreme Value Distributions: Example

Suppose that fissure lengths, X, in a rock mass have an exponential distribution with $f_X\left(x\right)=e^{-x}$. What, then, does the distribution of the maximum fissure length, Y_n, look like for n = 1, 5, and 50 fissures?

Solution:

If n = 1, then Y_n is the maximum of one observed fissure, which of course is just the distribution of the single fissure length. Thus, when n = 1, the distribution of Y_n is just the exponential distribution;

$$f_{Y_1}\left(y\right)=f_X\left(y\right)=e^{-y}$$

Exact Extreme Value Distributions: Example

When $n = 5$, we have

$$F_{Y_5}(y) = \mathrm{P}[Y_5 \le y] = \mathrm{P}[X_1 \le y]\mathrm{P}[X_2 \le y]\cdots\mathrm{P}[X_5 \le y] = [F_X(y)]^5$$

$$= [1 - e^{-y}]^5$$

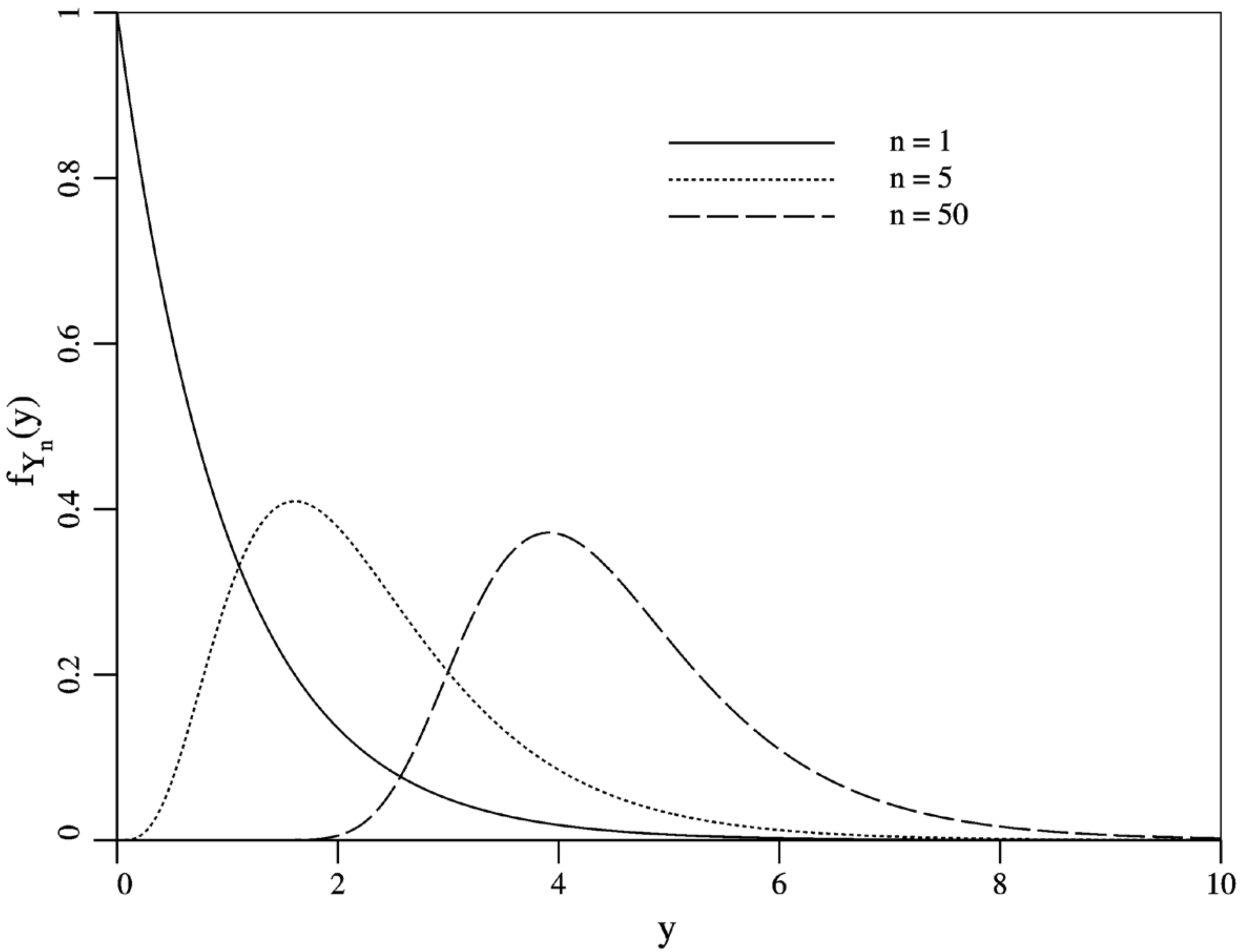
$$f_{Y_5}(y) = \frac{d\,F_{Y_5}(y)}{dy} = 5e^{-y}[1 - e^{-y}]^4$$

Similarly, when $n = 50$, we have

$$F_{Y_{50}}(y) = \mathrm{P}[Y_{50} \le y] = [F_X(y)]^{50} = [1 - e^{-y}]^{50}$$

$$f_{Y_{50}}(y) = \frac{d\,F_{Y_{50}}(y)}{dy} = 50e^{-y}[1 - e^{-y}]^{49}$$

Exact Extreme Value Distributions: Example



Exact Extreme Value Distributions: Minimum

Since Y_1 is the minimum of the X's, then the event $Y_1 > y$ implies that all of the X's are also greater than y. In other words,

$$\mathrm{P}\left[Y_1 > y\right] = \mathrm{P}\left[X_1 > y \cap X_2 > y \cap \cdots \cap X_n > y\right]$$

If it is assumed that the X's are independent and identically distributed, then

$$\mathrm{P}\left[Y_1 > y\right] = \mathrm{P}\left[X_1 > y\right]\mathrm{P}\left[X_2 > y\right]\cdots\mathrm{P}\left[X_n > y\right] = \left[1 - F_X(y)\right]^n$$

so that $F_{Y_1}(y) = 1 - \left[1 - F_X(y)\right]^n$

$$f_{Y_1}(y) = \frac{dF_{Y_1}(y)}{dy} = n\, f_X(y)\left[1 - F_X(y)\right]^{n-1}$$

Exact Extreme Value Distributions: Example

A series of 5 soil samples are taken at a site and their shear strengths determined. Suppose that a subsequent design is going to be based on the minimum shear strength observed out of the 5 samples. If the shear strengths of the individual samples are exponentially distributed with parameter $\lambda = 0.025 \ \text{m}^2/\text{kN}$ then what is the distribution of the design shear strength?

Solution:

$$F_{Y_1}(y) = 1 - \left[1 - F_X(y)\right]^5 = 1 - \left[1 - \left(1 - e^{-\lambda y}\right)\right]^5 = 1 - \left[e^{-\lambda y}\right]^5$$

$$= 1 - e^{-5\lambda y}$$

which is also exponential with 'rate' $n\lambda$