Chapter 3: Continuous Random Variables

So far, we've focused on random variables suitable for counts. The range of the random variables has been $\mathbb{N}$ or $\mathbb{Z}^+$.

We may also want to model phenomena taking on values in $\mathbb{R}$, for example time to completion, time to failure, inter-arrival time, and generally values relating to duration.

**Def.** A random variable $X$ on $(S, \mathcal{F}, P)$ is a function $X : S \rightarrow \mathbb{R}$ such that

$$\forall s \in S, \exists x_3 \in \mathcal{F} \text{ such that } x_3 = x \text{ for all } x \in \mathbb{R},$$

**Ex.** If $S = \mathbb{R}$ and $\mathcal{F}$ is a $\sigma$-field containing all open intervals $(a, b)$, then any piecewise continuous function $X : S \rightarrow \mathbb{R}$ is a valid random variable on $(S, \mathcal{F}, P)$

**Def.** The cumulative distribution function, CDF, $F_X$, associated with the random variable $X$ is the function

$$F_X : \mathbb{R} \rightarrow [0, 1] \text{ satisfying }$$

$$F_X(x) = P(X \leq x) \quad -\infty < x < \infty$$
(Note that this is the same as the definition applied to discrete random variables.)

Def. A **continuous random variable** $X$ is a random variable for which the CDF is continuous.

We will not show this, but it is true that, for a continuous random variable $X$, $F_X$ will be differentiable `almost everywhere', i.e. the set on which $\frac{dF}{dx}$ is undefined, say $D$ has

$P(x \in D) = 0$.

Def. The **probability density function**, pdf of a continuous random variable $X$ is defined by

$f(x) = \frac{dF}{dx}(x)$.

Note $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t) \, dt$

$\int_{-\infty}^{\infty} f(t) \, dt = 1$

$0 \leq F(x) \leq 1$
\[ \lim_{x \to -\infty} F(x) = 0 \quad \lim_{x \to \infty} F(x) = 1 \]

\[ P(X = c) = \int_c^\infty f(t) \, dt = 0 \]

\( F \) is monotone increasing, i.e. \( a > b \Rightarrow F(a) > F(b) \).

\( f(x) \geq 0 \) where defined.

\[ P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x) \, dx \]

for all continuous random variables \( X \) with CDF \( F \) and pdf \( f \).

Conversely, if \( f \) is a measurable function mapping \( \mathbb{R} \) to the non-negative reals, satisfying

\[ \int_{-\infty}^{\infty} f(t) \, dt = 1 \]

then \( f \) and \( F \) defined by

\[ F(x) = \int_{-\infty}^{x} f(t) \, dt \]

are the pdf and CDF of a continuous random variable. In particular, take \( S = \mathbb{R} \), \( f \) the smallest \( \sigma \)-field containing all open intervals, and define \( P \) by

\[ P(B) = \int_{t \in B} f(t) \, dt. \]
variable $X: S \to \mathbb{R}$ by $X(s) = s$ has $f$ and $F$ as its pdf and CDF, respectively.

Ex. Let $X$ be a random variable that takes on values in $[0, \frac{1}{2}]$ only, and takes those on with equal probability. The form of the pdf $f$ must be

$$f(x) = \begin{cases} 
  c & x \in [0, \frac{1}{2}] \\
  0 & \text{otherwise}
\end{cases}$$

(why?)

To pin this down, we know

$$1 = \int_{-\infty}^{\infty} f(t) \, dt = \int_{0}^{\frac{1}{2}} c \, dt = c \left[ t \right]_{0}^{\frac{1}{2}} = \frac{1}{2} c,$$

so $c = 2$.

The corresponding CDF $F$ is

$$F(x) = \begin{cases} 
  0 & x \leq 0 \\
  2x & 0 \leq x \leq \frac{1}{2} \\
  1 & x \geq \frac{1}{2}
\end{cases}$$
Graphically,

Recall from integral calculus that \( P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(t) \, dt \),
which represents the area under the function \( f \) between \( a \) and \( b \).

In the example above, \( P(\frac{1}{4} \leq X \leq 1) = \)
\[
= 2 \left[ \frac{1}{2} - \frac{1}{4} \right] \\
= \frac{1}{2}
\]

Can you think of a duration that could be modeled with a random variable like this one?
Most random variables we consider will be continuous or discrete.

The only remaining possibility is that $F$ is continuous except on a countable set on which it has jump discontinuities. (The proof of this is outside the scope of this course.)

Ex. Model the time to failure of a component that has an immediately fatal manufacturing flaw with probability 0.1. If the component does not have this flaw, then the probability that the component lasts until time $x$ drops off exponentially with $x$, i.e.

$$1 - F(x) = e^{-\lambda x} \quad \text{for } x > 0.$$ 

$$F(x) = 0 \quad x < 0 \quad (\text{easy})$$

$$F(0) = 0.1 \quad (\text{given})$$

$$F(x) = 1 - c e^{-\lambda x} \quad (\text{given})$$

The only discontinuity is at $x = 0$, and $\lim_{x \to 0^+} F(x) = 0.1$, so $c = 0.9$. 

and

\[ F(x) = \begin{cases} 0 & x < 0 \\ 0.1 & x = 0 \\ 1 - 0.9e^{-\lambda x} & x > 0 \end{cases} \]

\[ f(x) = \begin{cases} 0 & x < 0 \\ \text{undefined} & x = 0 \\ 0.9\lambda e^{-\lambda x} & x > 0 \end{cases} \]

We say that \( f \) has a point mass of mass 0.1 at \( x = 0 \). The value of \( \lambda \) would have to be determined experimentally.

\[ F(x) = \int \limits_{0}^{x} 0.9\lambda e^{-\lambda t} \, dt + 0.1 \]

\( C = \left[-0.9e^{-\lambda t}\right]_{0}^{x} + 0.1 = 0.9 - 0.9e^{-\lambda x} + 0.1 = 1 - 0.9e^{-\lambda x} \)
3.2 The Exponential Distribution

The exponential distribution is often used to model interarrival times of jobs, service time of jobs, time to failure, and time to repair. The memoryless property makes separable networks analytically manageable.

Exponential functions are a 1 parameter family of continuous random variables.

The CDF of the exponential distribution with parameter \( \lambda \) is \((\lambda > 0)\)

\[
F(x) = \begin{cases} 
1 - e^{-\lambda x} & 0 \leq x < \infty \\
0 & \text{otherwise}
\end{cases}
\]

with pdf

\[
f(x) = \begin{cases} 
\lambda e^{-\lambda x} & x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]
A random variable with this CDF and pdf is said to have the \( \text{EXP}(\lambda) \) distribution, \( X \sim \text{EXP}(\lambda) \).

The memoryless or Markov property of \( \text{EXP}(\lambda) \) is stated analytically as:

\[
P(X \leq y+t | X > t) = P(X \leq y).
\]

Pf. \( \frac{1-e^{-\lambda(y+t)}-(1-e^{-\lambda t})}{1-(1-e^{-\lambda t})} \)

\[
= \frac{e^{-\lambda t}(1-e^{-\lambda y})}{e^{-\lambda t}}
\]

\[
= 1-e^{-\lambda y}
\]

In fact, any nonnegative continuous random variable that is memoryless, i.e., satisfies \( P(X \leq y+t | X > t) = P(X \leq y) \), is exponential. The proof is in the text.

Why is a distribution satisfying \( \lambda \) called memoryless?
What are the implications of modeling time to failure with a memoryless distribution?

Important Relationship between Exponential RV and Poisson RV:

Interarrival times of Poisson RVs are exponentially distributed.

To gain intuition into why this is true, consider the time $X$ to the first arrival if the number of arrivals in $[0,t]$, $N_t$, is Poisson with parameter $\lambda t$.

$$P(X > t) = P(N_t = 0) = e^{-\lambda t} \frac{(\lambda t)^0}{0!}$$

$$= e^{-\lambda t}, \text{ so}$$

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

One can show that the numbers of arrivals over disjoint intervals are independent Poisson random variables, so this argument generalizes to interarrival times.
Ex. Jobs arrive at a CPU at an average rate of $\lambda = .05$ jobs/sec. Assuming that the number of arrivals in any time interval is Poisson distributed, what is the distribution of interarrival times? What is the probability of an interarrival time of $< .25$ sec?

$X \sim \text{EXP}(.05)$, so

$$F_x(.25) = 1 - e^{-.05(.25)} \approx .012$$

Try #1 §3.2 p. 123

1. Suppose CPU demand of jobs can be modeled as $\text{EXP}(\frac{1}{100\text{ms}})$. A job not completing in 100 ms is requeued. What is the probability that an arriving job will be requeued? Of 800 jobs, how many are expected to finish in the first quantum (100 ms)?
3.3 The Reliability, Failure Density, and Hazard Function

Let \( X \) be the lifetime or time to failure of a component. The probability that the component functions at least until time \( t \), \( P(X \geq t) = 1 - F(t) \), is called the reliability, \( R(t) \).

Compare the probability that the component will fail in \( [t, t+\Delta t] \), approximately \( f(t)\Delta t \), to the conditional probability that the component fails in \( (t, t+\Delta t) \) given that it functioned up to time \( t \):

\[
\frac{F(t+\Delta t) - F(t)}{1 - F(t)}
\]

Note

\[
\lim_{x \to 0} \frac{1}{x} \left( \frac{F(t+x) - F(t)}{1 - F(t)} \right) = \lim_{x \to 0} \frac{F(t+x) - F(t)}{x} \cdot \frac{1}{R(t)}
\]

\[= \frac{f(t)}{R(t)}\]

Def: The instantaneous failure rate,

\[ h(t) = \frac{f(t)}{R(t)} \]
This is also called the hazard rate, force of mortality, conditional failure rate, or failure rate.

Thus \( h(t)\Delta t \) is approximately the probability that the component will fail in \( (t, t+\Delta t] \) given that the component functions until time \( t \).

Exercise: Compare \( f(t) \) and \( h(t) \) for \( X(t) \) representing the lifetime of a human.

Ex. For \( F \sim \text{EXP}(\lambda) \), \( h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t} - 1} = \lambda \).

If you're given \( h \), you can reconstruct \( R \).

\[
R(t) = \exp \left[ -\int_0^t h(x) \, dx \right]
\]

Verification: \[
\int_0^t h(x) \, dx = \int_0^t \frac{f(x)}{R(x)} \, dx
\]
\[
= -\int_0^t \frac{R'(x)}{R(x)} \, dx \quad \text{change of variables } R \, (R(x)) \, dR = R' \, dx,
\]
\[
= -\int_{R(0)}^{R(t)} \frac{dR}{R} = -\ln(R(t)) - \ln(R(0))
\]

so \( R(t) = \exp \left[ -\int_0^t h(x) \, dx \right] \).
Def. The cumulative hazard, $H(t)$, is defined by

$$H(t) = \int_0^t h(x) \, dx$$

Thus $R(t) = e^{-H(t)}$

Note $h(t)$ constant $\iff$ $F(t)$ exponential

Exercise: Will a component that deteriorates as it ages have an increasing or decreasing hazard rate?

Can you think of a circumstance in which a component's conditional reliability improves with age?

What sort of situation is described by a bathtub shape for $h$?

Try problem 1, p. 129 §3.3
3.4 Some important distributions

Hypexponential Distribution: We will see that a sum of exponential random variables with distinct parameters \( \lambda_1, \lambda_2, \ldots, \lambda_r \) has a hypexponential distribution. This type of distribution has been shown to be a useful model for service times for IO operations.

To express \( F(t) \) and \( f(t) \) set

\[
a_i = \sum_{j=1}^{r} \frac{\lambda_j}{\lambda_j - \lambda_i} \quad \text{for} \quad i \neq j
\]

\[
\sum_{i=1}^{r} a_i = 1, \quad \text{though we won't prove this}
\]

Then

\[
f(t) = \sum_{i=1}^{r} a_i \lambda_i e^{-\lambda_i t}
\]

\[
F(t) = 1 - \sum_{i=1}^{r} a_i e^{-\lambda_i t} \quad \text{(add to p. 778)}
\]

For \( k=2 \),

\[
F(t) = 1 - \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 t}
\]

\[
f(t) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 t} - e^{-\lambda_2 t})
\]

\[
h(t) = \frac{\lambda_1 \lambda_2 (e^{-\lambda_1 t} - e^{-\lambda_2 t})}{\lambda_2 e^{-\lambda_1 t} - \lambda_1 e^{-\lambda_2 t}}
\]

\[
h(0) = 0, \quad h \text{ is increasing, } \lim_{t \to \infty} h(t) = \min \{ \lambda_1, \lambda_2 \}
\]
Erlang Distribution: We will later show this distribution is distribution of a sum of iid exponential distributions.

The pdf of the $r$-stage Erlang distribution with parameter $\lambda$ is

$$f(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!}, \quad t \in \mathbb{R}^+$$

$$F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad (p.777)$$

(obtained by repeated integration by parts)

Exercise: What random variable is a 1-stage Erlang RV?

The $r$-stage Erlang distribution models the time until the $r^{th}$ occurrence of phenomena Poisson distributed with parameter $\lambda$.

Gamma Distribution:

$$f(t) = \frac{\lambda^x t^{x-1} e^{-\lambda t}}{\Gamma(x)}, \quad t > 0, \lambda > 0$$

where $\Gamma(x)$ is the constant needed to make $f$ a pdf:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$
Note $\Gamma (1) = \int_0^\infty e^{-t} dt = 1$ and for $x > 1$

$$\int_0^\infty t^{x-1} e^{-t} dt = \left. -t^{x-1} e^{-t} \right|_0^\infty + \int_0^\infty (x-1) t^{x-2} e^{-t} dt = (x-1) \int_0^\infty t^{x-2} e^{-t} dt.$$ 

Thus $\Gamma (x) = (x-1) \Gamma (x-1)$.

Conclude $\Gamma (n) = (n-1)! \quad n \in \mathbb{Z}^+$

Also $\Gamma (1/2) = \sqrt{\pi}$, so $\Gamma (3/2) = \frac{1}{2} \sqrt{\pi}$,

$\Gamma (5/2) = \frac{3}{2} \cdot \frac{1}{2} \sqrt{\pi}$, etc.

$\Gamma (1/2, 1/2)$ is the $\chi^2$ distribution with $n$ degrees of freedom, useful in statistical analyses.

Hyperexponential Distribution: motivation---you have $k$ possible processes, each exponentially distributed with parameter $\lambda_i$, $i \in 1, 2, \ldots, k$, and selected with probability $\alpha_i$.

$$\sum_{i=1}^k \alpha_i = 1$$

$$f(t) = \sum_{i=1}^k \alpha_i \lambda_i e^{-\lambda_i t}$$

$$F(t) = \sum_{i=1}^k \alpha_i (1 - e^{-\lambda_i t})$$
Weibull: generalizes the exponential family of distributions. It offers good flexibility on the shape of the failure rate function.

\[ f(t) = \lambda t^{\alpha-1} e^{-\lambda t^\alpha} \]

\[ F(t) = 1 - e^{-\lambda t^\alpha} \]

\[ h(t) = \lambda t^{\alpha-1} \]

Log-logistic and Pareto distributions also described.

Normal or Gaussian Distribution

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right) \]

\[-\infty \leq x \leq \infty \]

\[ \sigma, \mu \in \mathbb{R} \]

\[ \sigma > 0 \]

Write \( X \sim N(\mu, \sigma^2) \)

\( F(x) \) has no closed form. The standard normal table on p. 789 is for \( Z \sim N(0,1) \) and shows

\[ F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} t^2 \right) dt \]

(Note the typos in the left-hand column. They should proceed in steps of .1: 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7... )
By symmetry, \( F_z(-z) = 1 - F_z(z) \).

For \( X \sim N(\mu, \sigma^2) \), \( \frac{X-\mu}{\sigma} \sim N(0,1) \)

\[ F_x(x) = F_z\left( \frac{x-\mu}{\sigma} \right) \]

as follows:

\[
\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2 \right) dt =
\]

(setting \( v = \left( \frac{t-\mu}{\sigma} \right) \), \( dv = \frac{1}{\sigma} \) \( dt \)

\( t=x \Rightarrow v = \frac{x-\mu}{\sigma} \))

\[
\int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (v)^2 \right) dv =
\]

\[ F_z\left( \frac{x-\mu}{\sigma} \right) . \]

ie: \( P\left( \frac{X-\mu}{\sigma} < t \right) = P\left( X < \sigma t + \mu \right) \)

\( = P\left( Z < \frac{(\sigma t + \mu) - \mu}{\sigma} \right) = P\left( Z < t \right) \)

Ex. A variety of battery is manufactured by a process that makes batteries with voltages normally distributed with \( \mu = 1.5 \) \( V \) and \( \sigma = .075 \) \( V \). What is the probability that a particular battery actually delivers 1.6 \( V \) or more?
\[ 1 - F_z \left( \frac{1.6 - 1.5}{0.75} \right) = 1 - F_z (1.3) \]
\[ \approx 1 - F_z (1.33) \approx 1 - 0.9082 = .0918 \]

using the normal table.

(The error function, erf, is available in some math packages.

\[ \text{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2} \, dy \]

This is useful for Normal calculations because, tracking down the integrals,

\[ \Phi_2 (1 + \text{erf} \left( \frac{z}{\sqrt{2}} \right)) = F_z (z) \]

In R, the CDF of \( N(\mu, \sigma) \) is \( \text{pnorm}(x, \mu, \sigma) \).

For intuition about \( N \), note
\[ P ( -1 \leq Z \leq 1 ) = 2 ( F_z (1) - \frac{1}{2}) = 2 \Phi_z (1) - 1 \]
\[ \approx .6826 \]

\[ P ( -2 \leq Z \leq 2 ) \approx .9542 \]
\[ P ( -3 \leq Z \leq 3 ) \approx .9974 \]
\[ P ( -4 \leq Z \leq 4 ) \approx 1.0000 \]
Uniform or Rectangular Distribution

\[ f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \]

\[ F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \]

Def. A defective distribution \( X \) is one for which

\[ \lim_{t \to \infty} F(t) < 1 \]

Try p. 147, #1 life times \( \sim N(5 \times 10^6, 5 \times 10^2 \mu) \)

Do you estimate that at least 95% have lifetimes greater than \( 4 \times 10^6 \) h?
Functions of a Random Variable

Given an RV $X$ and a function $\Phi : \text{image of } X \to \mathbb{R}$,

$Y = \Phi(X)$ will again be a random variable under fairly general conditions on $\Phi$, for example piecewise continuity.

You can often get $F_Y(y)$ from $F_X$ using inverse images:

$F_Y(y) = P(Y \leq y) = P(\exists x : \Phi(x) \leq y)$

Ex. $\Phi(X) = X^2$

$F_Y(y) = P(\exists x : x^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$.

Differentiate to get $f_Y(y)$:

$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) & y > 0 \\ 0 & \text{otherwise} \end{cases}$
Example 3.9:
In particular, if $X \sim N(0,1)$ and $Y = X^2$
then
$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, \quad y > 0$$
$$0 \quad \text{otherwise}$$

Check:

This makes $Y$ a Gamma distribution
with parameters $\frac{1}{2}$ and $\frac{1}{2}$, i.e. $\chi^2_1$.

Theorem: Let $X$ be a continuous random variable with density $f_x$. Define the set
$\mathcal{I} \subseteq \mathbb{R}$ by
$$\mathcal{I} = \{ x \in \mathbb{R} \mid f_x(x) > 0 \} \quad \text{(support of $f_x$)}$$
If $\Phi$ is a differentiable monotone function
$\Phi : \mathcal{I} \to \mathbb{R}$ and $Y = \Phi(X)$ then
$$f_Y(y) = \int f_x(\Phi^{-1}(y)) |(\Phi^{-1})' (y)| \quad y \in \Phi(\mathcal{I})$$
$$0 \quad \text{otherwise.}$$
Proof. \( f \) decreasing, i.e. \( y_0 \leq y \Rightarrow F(y_0) \geq F(y) \)

\[
F_Y(y) = P(Y \leq y) = P(F(X) \leq y)
\]

\[
= P(X \geq F^{-1}(y)) = 1 - F(F^{-1}(y))
\]

thus \( f_Y(y) = -f_X(F^{-1}(y))(F^{-1})'(y) \)

\[
= f_X(F^{-1}(y)) \frac{1}{(F^{-1})'(y)}
\]

An important consequence of this theorem is a method to simulate many RV's given a uniform distribution \( U \) on \( [0,1] \).

Given \( U \) uniform on \( [0,1] \) and an RV \( X \) which has \( F_X \) strictly monotone on \( I \), define \( \Phi : [0,1] \rightarrow \mathbb{R} \) by

\( \Phi = F_X^{-1} \). Set \( Y = \Phi(U) \). Then \( Y \) and \( X \) are identically distributed.

Proof. By the previous theorem,

\[
f_Y(y) = \begin{cases} f_U \left[ (F_X^{-1})'(y) \right] (F_X^{-1})'(y) & \text{for } y \in F_X^{-1}[0,1] \\ 0 & \text{otherwise} \end{cases}
\]
\[ f_X(y) = \int_0^1 f_X(y|x)f_X(x) \, dx \quad y \in \mathbb{R} \]

Otherwise:

\[ f_Y(y) = f_X \left( \frac{y-b}{a} \right) \frac{1}{a} \]

Example:

\[ Y = aX + b \quad \Phi^{-1}(y) = \frac{y-b}{a} \]

\[ \Phi^{-1}(y) = \frac{y-b}{a} \]

\[ f_Y(y) = f_X \left( \frac{y-b}{a} \right) \frac{1}{a} \]

Example:

\[ X \sim \text{Exp}(\lambda), \quad Y = cX, \quad f_Y(y) = \frac{1}{c} \lambda e^{-\lambda y}, \text{ so } Y \sim \text{Exp}(\lambda/c) \]

Example: Given a uniformly distributed RV \( U \) on \([0,1] \), what function of \( U \) is exponentially distributed with parameter \( \lambda \).

Set \( X \sim \text{Exp}(\lambda) \), \( F_X(x) = 1 - e^{-\lambda x} \quad x > 0 \)

Calculate \( F_X^{-1}(u) \)

\[ u = 1 - e^{-\lambda x} \quad e^{-\lambda x} = 1 - u \]

\[ \lambda x = \ln(1-u), \quad \text{so } F_X^{-1}(u) = \frac{\ln(1-u)}{-\lambda} \]

Conclude \( \frac{\ln(1-u)}{-\lambda} \sim \text{Exp}(\lambda) \)
Try p. 152 #3

$X$ uniform on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, $Y = \sin(X)$.

Calculate $f_y(y)$.

$\text{BTW } \arcsin'(y), \quad -\frac{\pi}{2} \leq \arcsin(y) \leq \frac{\pi}{2}$

is $\frac{1}{\sqrt{1-y^2}}$, as can be seen by applying the differentiation rule

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))} \quad \text{and} \quad \sin^2 \Theta + \cos^2 \Theta = 1$$
3.6 Jointly Distributed Random Variables

Recall that if $X, Y$ are RVs mapping $S \to \mathbb{R}$ then

$$F_{XY}(x,y) = P(X \leq x \text{ and } Y \leq y)$$

As in the discrete case,

$$0 \leq F(x,y) \leq 1 \quad x, y \in \mathbb{R}$$

$F(x,y)$ is monotone increasing in $x$ and $y$,

$$\lim_{x \to -\infty} F(x,y) = \lim_{y \to -\infty} F(x,y) = 0$$

$$\lim_{x,y \to \infty} F(x,y) = 1$$

Note $F(x,y)$ is continuous if $X$ and $Y$ are continuous.

As before,

$$P(a \leq X \leq b \text{ and } c \leq Y \leq d) =$$

$$F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

\begin{tikzpicture}
  \draw[step=1cm,gray,very thin] (0,0) grid (3,3);
  \draw[->,thick] (-0.5,0) -- (3.5,0) node[right] {$x$};
  \draw[->,thick] (0,-0.5) -- (0,3.5) node[above] {$y$};
  \draw (0,0) rectangle (3,3);
  \draw (0,2) -- (3,2);
  \draw (1,0) -- (1,3);
  \draw (0,1) -- (3,1);
\end{tikzpicture}
The marginal distribution of $X$ is
\[ F_x(x) = \lim_{y \to \infty} F_{xy}(x,y), \]
likewise for $Y$.

**Def.** $X$ and $Y$ are **jointly continuous** if there exists a measurable function $f(x,y): \mathbb{R}^2 \to \mathbb{R}$ satisfying
\[ F_{xy}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v) \, dv \, du. \]

**Def.** $X$ and $Y$ are independent random variables if
\[ F_{xy}(x,y) = F_x(x)F_y(y), \quad x, y \in \mathbb{R}. \]

If $X$ and $Y$ are continuous, this is equivalent to
\[ f_{xy}(x,y) = f_x(x)f_y(y), \quad \text{a.e.} \quad x, y \in \mathbb{R}. \]

**Example:** Suppose the joint pdf of $X$ and $Y$ is
\[ f(x,y) = \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left( -\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)} \right), \quad |\rho|<1 \]
Find the marginal density $f_y$. 
Rewrite \( f(x,y) \) as

\[
\frac{1}{2\pi \sqrt{1-p^2}} \exp\left(- \left[ \frac{x^2 - 2p xy + p^2 y^2}{2(1-p^2)} + \frac{(1-p^2) y^2}{2(1-p^2)} \right] \right)
\]

So \( \int f(x,y) \, dx \)

\[
= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi (1-p^2)}} \exp\left(-\frac{(x-p y)^2}{2(1-p^2)}\right) \, dx
\]

this is the pdf for \( N(py, \sqrt{1-p^2}) \), so

equals 1

\[
= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)
\]

Thus \( Y \sim N(0,1) \)

Similarly, \( X \sim N(0,1) \)

Note \( X \) and \( Y \) are not independent.

Try #3 p. 157: \( f(x,y) = \frac{1}{200} \quad (x,y) \in A \quad \text{otherwise} \)

where \( A \) is the triangular region with vertices \((100,100), (100,120), \) and \((120,120)\)

\( X \) and \( Y \) represent the lifetimes of two series connected components. Find the reliability expression for the system.
3.7 Order Statistics

Let $X_1, X_2, \ldots, X_n$ be iid RV (this is called a random sample of $X$ with $X_i \sim X$).

Define $Y_1, Y_2, \ldots, Y_n$ to be a permutation of $X_1, X_2, \ldots, X_n$ satisfying

$$Y_1 \leq Y_2 \leq \ldots \leq Y_n$$

Then $Y_k$ is the $k^{th}$ order statistic of $X_1, X_2, \ldots, X_n$.
Theorem: Let \( X_1, X_2, \ldots, X_n \) be a random sample with CDF \( F \) for each \( X_i \). Then the CDF of \( Y_k \) is

\[
F_k(y) = \sum_{j=k}^{n} \binom{n}{j} (F(y))^j (1-F(y))^{n-j}
\]

because \( Y_k \leq y \iff \) at least \( k \) of the \( X_i \)'s lie in \((-\infty, y]\) (the \( X_i \)'s that permute to \( Y_1, \ldots, Y_k \))

so

\[
F_k(y) = \sum_{j=k}^{n} \binom{n}{j} (F(y))^j (1-F(y))^{n-j}
\]

In particular, \( F_{\min}(y) = F_1(y) = 1-(1-F(y))^n \)

\[
\begin{align*}
\{ & = 1-(\binom{n}{0} F(y)^0 (1-F(y))^n) = \text{theorem applied to } (F(y)+1-F(y))^n \\
& = \sum_{j=0}^{n} \binom{n}{j} F(y)^j (1-F(y))^{n-j} - (\binom{n}{0} F(y)^0 (1-F(y))^n) \\
& = \sum_{j=1}^{n} \binom{n}{j} F(y)^j (1-F(y))^{n-j}
\end{align*}
\]

\( F_{\max}(y) = F_n(y) = (F(y))^n \)
Note that if \( F \) is the CDF of the time to failure for each of \( n \) independent components,

\[
R_{\text{series}}(t) = 1 - F_{\min}(t) = 1 - (1 - (1 - F(t))^n) = (R(t))^n
\]

\[
R_{\text{parallel}}(t) = 1 - F_{\max}(t) = 1 - (F(t))^n = 1 - [1 - R(t)]^n
\]

In general, if \( R_i \) is the reliability of the \( i \)th independent component,

\[
R_{\text{series}}(t) = \prod_{i=1}^{n} R_i(t)
\]

\[
R_{\text{parallel}}(t) = 1 - \prod_{i=1}^{n} (1 - R_i(t))
\]

Example: Suppose the system of \( n \) components is parallel and \( F(t) = 1 - e^{-\lambda t}, R(t) = e^{-\lambda t} \)

\[
F_{\text{sys}}(t) = (1 - e^{-\lambda t})^n
\]

\[
f_{\text{sys}}(t) = \lambda n (1 - e^{-\lambda t})^{n-1} e^{-\lambda t}
\]

\[
h_{\text{sys}}(t) = \frac{f_{\text{sys}}(t)}{R_{\text{sys}}(t)} = \frac{\lambda n (1 - e^{-\lambda t})^{n-1} e^{-\lambda t}}{1 - (1 - e^{-\lambda t})^n}
\]

Note \( h_{\text{sys}}(t) \) depends on \( t \), though the individual \( h \)'s are constant.
The probability that the system functions until time $t$ is $P(X > t \text{ and } Y > t)$

For $t < 100$ \[ P(X > t \text{ and } Y > t) = 1 \]

For $t < 120$ \[ P(X > t \text{ and } Y > t) = 1 - \frac{1}{200} \cdot \frac{(t-100)^2}{2} \]

For $t > 120$ \[ P(X > t \text{ and } Y > t) = 0. \]
Try problem 2, p. 165: A parallel system has 3 independent components with times to failure distributed as \( \text{EXP}(\lambda_i) \) where \( \lambda = 0.0001, 0.0002, 0.0004 \).

a. Determine the probability that the system will work for 1,000 h.

3.8. Distribution of Sums

Suppose \( X \) and \( Y \) are jointly continuous random variables with joint density \( f(x, y) \). Suppose the random variable \( Z \) is a function of \( X \) and \( Y \), \( Z = \Phi(x, y) \). For \( z \in \mathbb{R} \), define

\[
A_z = \{ (x, y) : \Phi(x, y) \leq z \} = \Phi^{-1}((-\infty, z])
\]

Then the CDF for \( Z \) has the formula

\[
F_Z(z) = P(Z \leq z) = \iint_{A_z} f(x, y) \, dx \, dy
\]

The sum \( Z = X + Y \) is a frequently used random variable.
For $Z = X + Y$, the set $A_2$ is graphed here.

So $F_2(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{xy} dt \, dx$. Change variable by $y = t - x$, $t = y + x$ to get

$$F_2(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f(x, t-x) \, dx \, dt$$

Thus

$$f_2(z) = \int_{-\infty}^{\infty} f(x, z-x) \, dx$$

If $X$ and $Y$ are independent, this becomes

$$f_2(z) = \int_{-\infty}^{\infty} f(x) f_y(z-x) \, dx \quad -\infty < z < \infty$$

$$= \int_{-\infty}^{\infty} f_y(y) f_x(z-y) \, dy$$
Example: (3.25) Cold Spare: Two statistically identical components are available for a system. Only one is required for the system to function properly. Keep one component as a powered-off spare. Assume it does not fail in this state. Assume failure detection and switching are perfect. Run the system until the first component fails, then swap in the second. The lifetime $Z$ of the system is $X+Y$ where $X$ and $Y$ are the lifetimes of the components. Now assume $X$ and $Y$ are iid exponentials: $f_X(x) = \lambda e^{-\lambda x}, \ x \geq 0, \ 
abla f_Y(y) = \lambda e^{-\lambda y}, \ y \geq 0, \ \nabla$.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \quad z \geq 0$$

$$= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda (z-x)} = \int_0^z \lambda^2 e^{-\lambda x} \ dx$$

$$\lambda^2 e^{-\lambda x} \int_0^z = \lambda z e^{-\lambda z} \quad z \geq 0$$

Thus $Z$ has a 2-stage Erlang distribution.

$$F(t) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} \quad t \geq 0$$

$$R(t) = (1 + \lambda t) e^{-\lambda t}$$

This is a special case of

Thm. If $X_1, X_2, \ldots, X_r$ are iid, $X_i \sim \text{EXP}(\lambda) \ \forall i$, then $X_1 + X_2 + \ldots + X_r$ has the $r$-stage Erlang distribution with parameter $\lambda$. (Pf. in Ch. 4)
If $X_1, X_2$ independent, $X_1 \sim \text{EXP}(\lambda_1)$, $X_2 \sim \text{EXP}(\lambda_2)$, then $Z = X_1 + X_2$ has a 2-stage hypoexponential distribution $Z \sim \text{Hypo}(\lambda_1, \lambda_2)$.

Calculation:

\[
\mathbb{P}(Z \leq z) = \int_0^z \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 (z-x)} \, dx
\]

\[
= \lambda_1 \lambda_2 \int_0^z e^{-\lambda_2 x} (\lambda_2 - \lambda_1) x \, dx
\]

\[
= \lambda_1 \lambda_2 \frac{e^{-\lambda_2 x} (e^{\lambda_2 z} - \lambda_1 e^{\lambda_1 z})}{\lambda_2 - \lambda_1}
\]

\[
= \frac{x_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_2 x} (e^{\lambda_2 z} - e^{\lambda_1 z}) - \frac{x_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 x}
\]

\[
= \frac{x_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_2 x} + \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} e^{-\lambda_1 x}
\]

Again, this is a special case of a theorem we will not prove here.

Theorem: If $Z = \sum_{i=1}^\infty X_i$ where $X_1, X_2, \ldots, X_n$ are mutually independent, $X_i \sim \text{EXP}(\lambda_i)$, and the $\lambda_i$'s are distinct,

\[
f_Z(z) = \sum_{i=1}^\infty a_i \lambda_i e^{-\lambda_i z} \quad z \geq 0
\]
where \( a_i = \frac{r}{\prod_{j=1}^{r} \frac{\lambda_j}{\lambda_j - \lambda_i}} \)

\( Z \sim \text{HYPOCA}_1, \lambda_2, \ldots, \lambda_r \)

**Corollary:** the sum \( Z \) of two hypoequivalent distributions \( X_1 \sim \text{HYPOCA}_1, \lambda_2, \ldots, \lambda_r \), \( X_2 \sim \text{HYPOCA}_1, \lambda_2, \ldots, \lambda_r \), \( X_1, X_2 \) independent and \( \lambda_1, \lambda_2, \ldots, \lambda_r \) distinct has

\( Z \sim \text{HYPOCA}_1, \lambda_2, \ldots, \lambda_r \)  
(Why?)

**Example 3.29**

\[ 
\text{Input} \xrightarrow{\text{functional unit}} \downarrow \text{output} \\
\quad \downarrow \text{detector} \Rightarrow \text{error signal} \\
\]

\( T \): time to failure of unit
\( C \): time to failure of detector
\( D \): time to detection of unit failure
(Detector failure is reported instantaneously.)

\( X \): time to indication of failure of functional unit or detector

\( Y \): time to first failure occurrence (functional unit or detector)
\[ X = \min E, T + D \]
\[ Y = \min E, C \]

Suppose \( T, C, \) and \( D \) are independent exponential distributions with parameters \( \lambda_T, \lambda_C, \) and \( \lambda_D \) respectively.

For \( Y, \)
\[ F_Y(y) = 1 - (1 - F_T(y))(1 - F_C(y)) = 1 - e^{-(\lambda_T + \lambda_C)y} \]
so \( Y \sim \text{EXP}(\lambda_T + \lambda_C) \)

For \( X, \) \( T + D \) is hypoexponentially distributed
\[ F_{T+D}(t) = 1 - \frac{\lambda_T}{\lambda_T - \lambda_D} e^{-\lambda_D t} - \frac{\lambda_D}{\lambda_D - \lambda_T} e^{-\lambda_T t} \]
so the apparent reliability is
\[ R_X(t) = P(X \geq t) = P(\min \{T+D, C\} \geq t) \]
\[ = P(T+D \geq t)P(C \geq t) \]
\[ = \left( \frac{\lambda_T}{\lambda_T - \lambda_D} e^{\lambda_D t} + \frac{\lambda_D}{\lambda_D - \lambda_T} e^{\lambda_T t} \right) e^{-\lambda_C t} \]

Consider a system with \( n \) iid components with failure law \( \text{EXP} (\lambda). \) If \( k \) or more components must be functioning for the system to function then the lifetime of the system is given by the order statistic \( Y_{n-k+1} \).

\( (n-k+1) \) components have failed \( \Rightarrow n - (n-k+1) = k-1 \) are working, i.e. the system just failed.
Theorem 3.5  If \( X_1, X_2, \ldots, X_n \) are iid with \( X_i \sim \text{EXP}(\lambda) \) then

\[
Y_{n-k+1} \sim \text{HYPO} [n\lambda, (n-1)\lambda, \ldots, k\lambda],
\]

an \( n-k+1 \) stage hypoexponential distribution.

Pf. Proceed by induction on \( n-k+1 \).

Base case: \( n-k+1=1 \), i.e. \( n=k \).

\[
Y_1 = \min (X_1, X_2, \ldots, X_n) \sim \text{EXP}(n\lambda),
\]
as required.

Assume \( Y_{n-j+1} \sim \text{HYPO} (n\lambda, (n-1)\lambda, \ldots, j\lambda) \).

Show \( Y_{n-(j-1)+1} \sim \text{HYPO} (n\lambda, (n-1)\lambda, \ldots, (j-1)\lambda) \) given the assumption:

\[
Y_{n-j+2} = Y_{n-j+1} + \min (W_{n-j+2}, \ldots, W_n)
\]
where the \( W_i \)'s are the lifetimes of the surviving components. By the memoryless property, \( W_i \sim \text{EXP}(\lambda) \).

There are \( j-1 \) \( W_i \)'s, so the minimum is distributed as \( \text{EXP}((j-1)\lambda) \), a 1 stage hypoexponential distribution.

We have previously shown that a sum of hypoexponential distributions with distinct parameters is hypoexponentially distributed with the set of parameters equal to the union of the parameters of the summands.

Conclude \( Y_{n-j+2} \sim \text{HYPO} (n\lambda, (n-1)\lambda, \ldots, j\lambda, (j-1)\lambda) \).
A simplification of the coefficients appears on p. 179.

The following diagram shows another way to think of this result:

\[
\begin{align*}
\text{EXP}(n\lambda) + \text{EXP}((n-1)\lambda) + \cdots + \text{EXP}(k\lambda) + \cdots + \text{EXP}(\lambda) \\
Y_1 \quad Y_2 \quad Y_{n-k+1} \quad Y(n)
\end{align*}
\]

Example 3.32 p. 179 similarly applies induction and the memoryless property of exponential distributions to compute the lifetime of a \( k \)-out-of-\( n \) system with \( m \) spares.

Suppose we have a \( k \)-out-of-\( n \) system with \( n \) components and \( m \) spares. Active components' lifetimes are i.i.d. \( \text{EXP}(\lambda) \), while inactive components independently have time to failure \( \sim \text{EXP}(\mu) \). Then system lifetime, \( L(k|n,m) \), satisfies

\[
L(k|n,m) \sim \text{HYPO}\left(n\lambda + \mu, n\lambda + 2\mu, \ldots, n\lambda + m\mu, k\lambda, (k+1)\lambda, \ldots, n\lambda\right)
\]

The corresponding diagram,

\[
\begin{align*}
\text{EXP}(n\lambda + m\mu) \rightarrow \text{EXP}(n\lambda + (m-1)\mu) \rightarrow \cdots \text{EXP}(n\lambda + \mu) \rightarrow \text{EXP}(n\lambda) \rightarrow \\
\text{EXP}(n-k\lambda) \rightarrow \cdots \text{EXP}(k\lambda)
\end{align*}
\]
reflects the fact that, while there are \( p \) spares, the next component failure occurs at the minimum of \( n \) \( \text{EXP}(\lambda_i) \) RVs and \( p \) \( \text{EXP}(\mu_i) \) RVs. The system then has \( n \) active components and \( p-1 \) spares.

Once all the spares have been exhausted, the system is \( k \) out of \( n \) system, previously analyzed.

Try problem 1, p. 182: Suppose you have a source of \( n \) random values \( \mu_1, \ldots, \mu_n \), uniformly distributed on \([0,1] \). Give a formula in the \( \mu_i \)'s that results in a random variable that is hypoexponentially distributed with parameters \( \lambda_1, \lambda_2, \ldots, \lambda_n \).

(Hints: If \( U \) is uniform on \([0,1] \), \( F_X^{-1}(U) \sim X \).
If you had \( X_i \sim \text{EXP}(\lambda_i) \) \( i = 1, \ldots, n \), how could you generate a \( Z \sim \text{HYPO}(\lambda_1, \lambda_2, \ldots, \lambda_n) \)?)
3.9 Functions of Normal Random Variables

Normal distributions are common approximating distributions in statistics due to the Central Limit Theorem. As a result, distributions of functions of Normal RVs acquire importance. Some results are given below.

**Theorem 3.6** \( X_1, X_2, \ldots, X_n \) mutually independent \( X_i \sim N(\mu_i, \sigma_i^2) \) then \( S_n = \sum_{i=1}^{n} X_i \) satisfies

\[
S_n \sim N\left( \sum \mu_i, \sum \sigma_i^2 \right)
\]

**Proposition** \( X \sim N(\mu, \sigma^2) \) then \( aX \sim N(a\mu, a^2\sigma^2) \)

Can you show this based on \( F_{aX}(t) = F_X \left( \frac{t}{a} \right) \)?

so \( f_{aX}(x) = \frac{1}{a} f_X \left( \frac{x}{a} \right) \).
Example 3.34: Suppose $X_1, X_2, \ldots, X_n$ iid, $X_i \sim N(\mu, \sigma^2)$. Set $X = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$

Then $\overline{X} \sim N\left(\frac{\sum_{i=1}^{n} \mu_i}{n}, \frac{\sum_{i=1}^{n} \sigma^2}{n^2}\right) = N(\mu, \frac{\sigma^2}{n})$

The distribution of the mean $X$ has the same center as the $X_i$'s and a sharper peak. This is why the mean of normally distributed measurements centered on some value $\mu$ is a better estimate of $\mu$ than a single measurement.

Sums of Squares of iid $X_1, X_2, \ldots, X_n: X_i \sim N(0, 1)$

If $X \sim N(0, 1)$ then $Y = X^2$ satisfies

$\quad Y \sim \text{GAM}(\frac{1}{2}, \frac{1}{2}) = X_1$

($X_1 = \text{GAM}(\frac{1}{2}, \frac{n}{2})$)

We showed this as Example 3.9.

Example 3.35: If $X_1, X_2$ iid, $X_i \sim N(0, 1)$
then $Y = X_1^2 + X_2^2$ satisfies

$\quad Y \sim \text{EXP}(\frac{1}{2}) \cdot \text{GAM}(\frac{1}{2}, 1) = X_2$

This is proved by hand in the text.
More generally

**Theorem 3.7.** If $X_1, X_2, \ldots, X_n$ iid standard normal and $Y = \sum_{i=1}^{n} X_i^2$ implies

$$Y \sim \text{GAM}(\frac{n}{2}, \frac{1}{2}) = \chi_n^2$$

This is a consequence of

**Theorem 3.8.** Let $X_1, \ldots, X_n$ be a sequence of mutually independent, random variables with

$$X_i \sim \text{GAM}(\lambda, \alpha_i).$$

Set $S_n = \sum_{i=1}^{n} X_i$. Then

$$S_n \sim \text{GAM}(\lambda, \sum_{i=1}^{n} \alpha_i)$$

**Application (glimpse of Ch. 10).** If you know $X_1, \ldots, X_n$ iid, $X_i \sim N(\mu, \sigma^2)$ with $\mu$ known and $\sigma^2$ unknown and you want to approximate $\sigma^2$ from values of the $X_i$'s (aka 'data'), note

$$\sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2.$$  If we use

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 = S^2,$$  say, as an estimator for $\sigma^2$, we can obtain a confidence
interval for our result.

For example, with \( n=10 \), and a 95\(^{th} \) confidence interval required, use the \( X^2 \) table to find values \( a \) and \( b \) such that \( X \sim \chi^2_{10} \) implies

\[
P(a < X < b) = .95.
\]

If we pick \( a = 3.24697 \) and \( b = 20.4832 \)

\[
P(a \leq X^2_{10} \leq b) = .975
\]

\[
P(X^2_{10} \leq b) = .025
\]

so

\[
P\left( a \leq X^2_{10} \leq b \right) = .95
\]

Then

\[
P\left( \frac{10S^2}{a} \leq \frac{10S^2}{b} \leq b \right) = .95
\]

\[
P\left( \frac{1}{b} \leq \frac{\sigma^2}{10S^2} \leq \frac{1}{a} \right) = .95
\]

\[
P\left( \frac{10S^2}{b} \leq \sigma^2 \leq \frac{10S^2}{a} \right) = .95
\]

so

\[
P\left( .49S^2 \leq \sigma^2 \leq 3.08S^2 \right) = .95
\]

\[
P\left( .7S \leq \sigma \leq 1.8S \right) = .95
\]
Now we are prepared to say what we mean by a 95% confidence interval: if we calculate the upper and lower bounds from the RVs $X_i$, the probability that the interval will include the value of $\sigma^2$ is .95.

Actually, often $\mu$ and $\sigma^2$ are unknown. Then the estimator

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

is used.

It turns out that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$,

making the calculations for the confidence intervals for this estimator very similar to those above.

**Theorem 3.9** Let $Y_1$ and $Y_2$ be independent RVs $Y_1 \sim \chi^2_{n_1}$, $Y_2 \sim \chi^2_{n_2}$ then $Z$ 

$$Z = \frac{Y_1/n_1}{Y_2/n_2}$$

has the $F_{n_1, n_2}$ distribution, for which tables are available.

Pf in text.
Theorem 3.10 If $V$ and $W$ are independent random variables such that $V \sim N(0,1)$ and $W \sim X_n^2$, then

$$T = \frac{V}{\sqrt{\frac{W}{n}}}$$

has the $t$-distribution with $n$ degrees of freedom.

The pdf of the $t$-distribution with $n$ degrees of freedom is

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left[1 + \frac{t^2}{n}\right]^{-(n+1)/2}$$

As $n \to \infty$, $f_{T_n}(t)$ approaches $f_2(t)$, the pdf of the standard normal random variable.

Application: This gives us an estimator for $\mu$ of a random sample $X_1, X_2, \ldots, X_n$, $X_i \sim N(\mu, \sigma^2)$, if both $\mu$ and $\sigma$ are unknown.

Recall $V = \frac{(X-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$. Also

$$W = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left[\frac{X_i - \overline{X}}{\sigma}\right]^2 \sim X_{n-1}^2$$

Thus

$$T = \frac{V}{\sqrt{\frac{W}{n-1}}} = \frac{(X-\mu)\sqrt{n}}{\sigma \sqrt{(n-1)S^2 / \sigma^2}} = \frac{X-\mu}{\frac{S}{\sqrt{n}}}$$

has the $t$-distribution with $n-1$ degrees of freedom.
To find a confidence interval for \( \mu \), use a t-table to find \( t \) such that

\[ 1 - F_{t_n}(a) = \frac{1}{2}, \quad \text{say.} \]

Then, by the symmetry of the t-distribution around 0,

\[ P(-a \leq T \leq a) = 1 - \alpha \]

\[ P(-a = \frac{\bar{x} - \mu}{s/n} \leq a) = 1 - \alpha \]

\[ P\left(-\frac{2s}{\sqrt{n}} \leq \bar{x} - \mu \leq \frac{2s}{\sqrt{n}}\right) = 1 - \alpha \]

\[ P\left(-\frac{2s}{\sqrt{n}} \leq -\mu \leq -\bar{x} + \frac{2s}{\sqrt{n}}\right) = 1 - \alpha \]

\[ P\left(\frac{2s}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{2s}{\sqrt{n}}\right) = 1 - \alpha \]

Try #3 53.9 p. 190.