STA 4033
Mathematical Statistics with Computer Applications
Lectures 5 through 7*

Brett Presnell
Department of Statistics
University of Florida

21 May 2001

*Copyright ©2001 by Brett Presnell
3. Random Variables and their Probability Distributions

A statistical experiment is a procedure whose outcome cannot be predicted with certainty, but for which the outcomes nevertheless have stable “long run relative frequencies,” so that probabilities can generally be assigned to events of interest. The probability of an event reflects either

- the relative frequency of the event in an infinite sequence of independent repetitions of the experiment (frequentist interpretation), or
- how likely we feel that the event is to occur (subjective interpretation).

We will generally interpret probability in frequentist terms.
3.1. Some Terminology

**Sample point:** one of the individual possible outcomes of an experiment.

**Sample space:** the collection of *all* sample points.

**Event:** a set of sample points.

**Random variable:** a real-valued function defined on the sample space.

**Distribution of a random variable:** the collection of all events that can be defined in terms of the random variable together with their associated probabilities. Usually summarized as a function or table sufficient to calculate all such probabilities.

**Distribution Function:** One way to describe the probability distribution of a random variable is to give its *(cumulative)* distribution function:

\[ F(x) = P(X \leq x). \]

**Mean of** \(X\): the long run average value of \(X\) in infinitely many independent repetitions of the experiment.

- Also called the *expected value* of \(X\).
- Denoted \(\mu, \mu_X\), and/or \(E(X)\).

**Variance of** \(X\): the average squared distance of \(X\) from its mean. Mathematically,

\[ \sigma^2 = \sigma_X^2 = \text{Var}(X) = E[(X - \mu)^2]. \]
3.1.1. An Example: Coin Tossing

Suppose that we toss a fair coin three times (tosses assumed independent), recording heads (H) or tails (T) for each toss. One random variable of interests might be $X$, the number of heads. Here is the sample space, the probability associated with each sample point, and the value of $X$ for each sample point.

<table>
<thead>
<tr>
<th>Sample Point</th>
<th>Probability</th>
<th>Value of X</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHH</td>
<td>1/8</td>
<td>3</td>
</tr>
<tr>
<td>HHT</td>
<td>1/8</td>
<td>2</td>
</tr>
<tr>
<td>HTH</td>
<td>1/8</td>
<td>2</td>
</tr>
<tr>
<td>HTT</td>
<td>1/8</td>
<td>1</td>
</tr>
<tr>
<td>THH</td>
<td>1/8</td>
<td>2</td>
</tr>
<tr>
<td>THT</td>
<td>1/8</td>
<td>1</td>
</tr>
<tr>
<td>TTH</td>
<td>1/8</td>
<td>1</td>
</tr>
<tr>
<td>TTT</td>
<td>1/8</td>
<td>0</td>
</tr>
</tbody>
</table>

The distribution of $X$ is given by the following table:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/8</td>
</tr>
<tr>
<td>1</td>
<td>3/8</td>
</tr>
<tr>
<td>2</td>
<td>3/8</td>
</tr>
<tr>
<td>3</td>
<td>1/8</td>
</tr>
</tbody>
</table>
### 3.2. Discrete Random Variables

A **discrete** random variable is one having only finitely or countably many possible values, i.e., the possible values can be enumerated.

- The most common examples are random variables that are counts, and thus take on nonnegative integer values.

- The distribution of a discrete random variable is usually given by its *probability function*,

\[
p(x) = P(X = x).
\]

- The mean and the variance are calculated via

\[
\mu = E(X) = \sum_x x p(x) \quad \text{and} \quad \text{Var}(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 p(x).
\]

- In general, the expectation of any function \(g(\cdot)\) of \(X\) is calculated as

\[
E[g(X)] = \sum_x g(x) p(x).
\]

We present some of the distributions commonly used to model the probabilistic behavior of discrete random variables.
3.2.1. The Binomial Distribution

A binomial experiment consists of

- a fixed number, \( n \), of
- independent trials
- with each trial resulting in success (S) or failure (F),
- with probability \( p \) of S and \( q = 1 - p \) of F constant from trial to trial.

In this case, the random variable

\[ X = \text{number of successes} \]

is said to have a binomial distribution on \( n \) trials with success probability \( p \).

- The probability function of \( X \) is

\[ p(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n. \]

- The mean and variance of \( X \) are

\[ \mu = np \quad \text{and} \quad \sigma^2 = npq. \]
• Note that the binomial distribution is actually a family of distributions.

• The distribution is not specified exactly unless we know $n$ and $p$.

• $n$ and $p$ are called parameters of the distribution.

Figure 1 shows a plot of the binomial probability function for $n = 10$ and $p = 0.2$, $0.5$, and $0.8$. Note that the distribution is

• skewed right when $p < 1/2$;

• skewed left when $p > 1/2$;

• symmetric when $p = 1/2$. 
Figure 1: Binomial distributions with $n = 10$ and $p = 0.2, 0.5, 0.8$. 
The following R functions are provided for the binomial distribution:

- **dbinom**: probability function (density).
- **pbinom**: distribution function.
- **qbinom**: quantile function.
- **rbinom**: generate binomial random variables.

The plots in Figure 1 were produced by the commands:

```r
> opar <- par(mfrow=c(1,3))
> plot(0:10, dbinom(0:10, 10, 0.2), type="h", xlab="", ylab="Probability ", sub="p=0.2")
> plot(0:10, dbinom(0:10, 10, 0.5), type="h", xlab="", ylab="Probability ", sub="p=0.5")
> plot(0:10, dbinom(0:10, 10, 0.8), type="h", xlab="", ylab="Probability ", sub="p=0.8")
```
3.2.2. The Hypergeometric Distribution

An urn contains $N$ balls, $r$ red and $N - r$ blue. We will draw $n$ balls from the urn and let $X$ be the number of red balls in our sample. Consider two sampling schemes:

**Sampling with replacement:** we draw $n$ balls at random, one at a time, replacing each ball before drawing the next.

- Then $X$ has a binomial distribution on $n$ trials with success probability $p = r/N$.

**Sampling without replacement:** we draw $n$ balls at random without replacement ($n \leq N$).

- Note that the draws are *not* independent.
- In this case $X$ has a *hypergeometric* distribution.
- The probability function of $X$ is

$$p(x) = \frac{r \choose x}{{N \choose n}} \frac{(N-r) \choose (n-x)}{{N \choose n}}, \quad 0 \leq x \leq n, \quad 0 \leq x \leq r, \quad n - x \leq N - r.$$

- Let $p = r/N$ and $q = 1 - p = (N - r)/N$. Then the mean and variance of $X$ are

$$\mu = np \quad \text{and} \quad \sigma^2 = npq \frac{N-n}{N-1}.$$
**Note:** When $N$ is large (and $n$ is moderate), then the hypergeometric distribution is practically indistinguishable from the binomial.

This is illustrated in Figure 2 where the sample size is $n = 5$, the fraction of red balls (successes) is $p = 0.4$, and the total number of balls in the urn is variously taken to be $N = 10, 20, 100,$ and $\infty$ (binomial).

Here is a table of the corresponding probabilities, with the case $N = 1000$ also included:

<table>
<thead>
<tr>
<th>$N$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.024</td>
<td>0.238</td>
<td>0.476</td>
<td>0.238</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>20</td>
<td>0.051</td>
<td>0.255</td>
<td>0.397</td>
<td>0.238</td>
<td>0.054</td>
<td>0.004</td>
</tr>
<tr>
<td>100</td>
<td>0.073</td>
<td>0.259</td>
<td>0.355</td>
<td>0.232</td>
<td>0.073</td>
<td>0.009</td>
</tr>
<tr>
<td>1000</td>
<td>0.077</td>
<td>0.259</td>
<td>0.346</td>
<td>0.231</td>
<td>0.076</td>
<td>0.010</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.078</td>
<td>0.259</td>
<td>0.346</td>
<td>0.230</td>
<td>0.077</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Figure 2: Hypergeometric distributions with $n = 5$, $r = 0.4N$, and $N = 10, 20, 100, \infty$ (binomial).
The following R functions are provided for the hypergeometric distribution:

\texttt{dhyper}: probability function (density).

\texttt{phyper}: distribution function.

\texttt{qhyper}: quantile function.

\texttt{rhyper}: generate hypergeometric random variables.

The plots in Figure 2 were produced by the commands

```r
> opar <- par(mfrow=c(2,2))
> N <- 10; plot(0:5, dhyper(0:5, 0.4*N, 0.6*N, 5), type="h",
  xlab="", ylab="Probability ", sub="N=10, r=4, n=5")
> N <- 20; plot(0:5, dhyper(0:5, 0.4*N, 0.6*N, 5), type="h",
  xlab="", ylab="Probability ", sub="N=20, r=8, n=5")
> N <- 100; plot(0:5, dhyper(0:5, 0.4*N, 0.6*N, 5), type="h",
  xlab="", ylab="Probability ", sub="N=100, r=40, n=5")
> plot(0:5, dbinom(0:5, 5, 0.4), type="h",
  xlab="", ylab="Probability ", sub="p=0.4, n=5")
> par(opar) # restore the default graphics setup
```
The earlier table was produced by the commands

```r
> tmp <- t(sapply(c(10,20,100,1000),
             function(N) (dhyper(0:5, 0.4*N, 0.6*N, 5)))))
> tmp <- rbind(tmp, dbinom(0:5, 5, 0.4))
> round(tmp, 3)
```
3.2.3. The Geometric Distribution

Let $X$ be the number of failures before the first success in a sequence of independent trials, each yielding a success with probability $p$ and failure with probability $q$. Then $X$ has a geometric distribution.

- The probability function of the geometric distribution is
  
  $$p(x) = q^x p, \quad x = 0, 1, 2, \ldots$$

- The mean and variance of the geometric distribution are
  
  $$\mu = \frac{1}{p} - 1 \quad \text{and} \quad \sigma^2 = \frac{q}{p^2}.$$ 

The following R functions are provided for the geometric distribution:

- `dgeom`: probability function (density).
- `pgeom`: distribution function.
- `qgeom`: quantile function.
- `rgeom`: generate geometric random variables.
3.2.4. The Negative Binomial Distribution

Let $X$ be the number of failures before the $r$th success in a sequence of independent trials, each yielding a success with probability $p$ and failure with probability $q$. Then $X$ has a negative binomial distribution.

- The probability function of the negative binomial distribution is

$$p(x) = \binom{x + r - 1}{r - 1} q^x p^r, \quad x = 0, 1, 2, \ldots$$

- The mean and variance of the geometric distribution are

$$\mu = r \left( \frac{1}{p} - 1 \right) \quad \text{and} \quad \sigma^2 = \frac{rq}{p^2}.$$

The following R functions are provided for the negative binomial distribution:

- `dnbinom`: probability function (density).
- `pnbinom`: distribution function.
- `qbinom`: quantile function.
- `rnbinom`: generate negative binomial random variables.
3.2.5. The Poisson Distribution

The Poisson distribution is not as easily motivated as the distributions above, but it is at least as important in applications. For example, the Poisson is often used to model counts of events occurring randomly in space or in time. Typical examples include:

- the number of alpha particles emitted by a quantity of uranium in a given time period.
- the number of firing’s of a neuron during a given time period.
- the number of accidents at a given intersection during a given time period.
- the number of packets arriving at a router during a given time period.
- the number of trees of a certain species within a given parcel of land.
- the number of flaws in a given length of fabric.
- the number of errors in a page of typing.
• The Poisson distribution with mean $\lambda$ has probability function

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0, 1, 2, \ldots.$$ 

• The variance of the Poisson distribution is equal to its mean, $\lambda$, i.e., for a Poisson distribution with parameter $\lambda$,

$$\mu = \sigma^2 = \lambda.$$ 

Note: When $n$ is very large and $p$ is very small, the binomial distribution can be approximated by a Poisson distribution with mean $\lambda = np$.

This is demonstrated in Figure 3, where binomial distributions with $n = 10$, 20, and 100 and with $p = 2/n$ are compared to the Poisson distribution with $\lambda = 2$. 
Figure 3: Binomial distributions with $n = 10, 20, \text{ and } 100$ and $p = \frac{2}{n}$ compared to Poisson with mean 2.
The following R functions are provided for the Poisson distribution:

- **dpois**: probability function (density).
- **ppois**: distribution function.
- **qpois**: quantile function.
- **rpois**: generate Poisson random variables.

The plots in Figure 3 were produced by the commands

```r
> opar <- par(mfrow=c(2,2))
> lambda <- 2
> for (n in c(10,20,100)) {
  p <- lambda/n
  tmp <- paste("n = ", n, ", p = ", p, sep=" ")
  plot(0:10, dbinom(0:10, n, lambda/n), type="h",
       xlab="", ylab="Probability ", sub=tmp)
}
> tmp <- paste("Poisson ( ", lambda, ") ", sep=" ")
> plot(0:10, dpois(0:10, lambda), type="h",
      xlab="", ylab="Probability ", sub=tmp)
> par(opar) # restore default graphics parameters
```
3.3. Continuous Random Variables

A continuous random variable is having the property that

\[ P(X = x) = 0 \quad \text{for all real numbers } x. \]

- Definition is counter-intuitive.
  - Think of measuring something (e.g., time to complete a task, height, etc.) with infinite precision.
  - In practice continuous distributions are also used to model measurements (like IQ and SAT scores) that are discrete but have very many possible values.

- Note that

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

- The distribution of a continuous random variable is usually given by its probability density function, denoted say \( f(x) \).

- Probabilities are calculated using the density via

\[ P(a < X < b) = \int_{a}^{b} f(x) \, dx. \]
• The mean and the variance are calculated via

\[ \mu = E(X) = \int_{-\infty}^{\infty} x f(x) \, dx \quad \text{and} \quad \text{Var}(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx. \]

• In general, the expectation of any function \( g(\cdot) \) of \( X \) is calculated as

\[ E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx. \]
3.3.1. The Uniform Distribution

A *uniform* distribution is one with a constant density over some region. Thus the uniform distribution on the interval \((\theta_1, \theta_2)\) has density

\[
f(x) = \frac{1}{\theta_2 - \theta_1}, \quad \theta_1 < x < \theta_2.
\]

- All values between \(\theta_1\) and \(\theta_2\) are equally likely.
- The uniform \((\theta_1, \theta_2)\) distribution has mean and variance given by

\[
\mu = \frac{\theta_1 + \theta_2}{2} \quad \text{and} \quad \sigma^2 = \frac{(\theta_2 - \theta_1)^2}{12}.
\]

Note that \(\mu\) is the midpoint of the interval and \(\sigma^2\) is the length of the interval squared and then divided by 12.

- *THE* uniform distribution is usually taken to be the uniform distribution on the unit interval \((0, 1)\).

The following R functions are provided for the uniform distribution:

- **dunif**: density function.
- **punif**: distribution function.
- **qunif**: quantile function.
- **runif**: generate uniform random variables.
3.3.2. The Normal Distribution

A normal distribution with mean $\mu$ and variance $\sigma^2$ has density

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}, \quad -\infty < x < \infty.$$ 

- This is the famous “bell curve” distribution.

- Also called the “Gaussian distribution” in honor of Karl Frederick Gauss who used it to model errors in astronomical observations.

- The standard normal distribution has mean 0 and variance 1.

- The normal distribution is used to model all sorts of things, including:
  
  - measurement error.
  - random “noise”.
  - the distribution of physical measurements such as heights and weights in a “homogeneous” population.
  - scores on standardized tests.

- The normal distribution also arises in many situations through the operation of the central limit theorem.
The following R functions are provided for the normal distribution:

- **`dnorm`**: density function.
- **`pnorm`**: distribution function.
- **`qnorm`**: quantile function.
- **`rnorm`**: generate normal random variables.

The plots in Figure 4 compare normal distributions with the same mean (0) but different standard deviations, and were produced by the commands:

```r
> x <- seq(-10, 10, length=100)
> plot(x, dnorm(x, 0, 1), type="l", xlab="x", ylab="f(x)", lty=1)
> lines(x, dnorm(x, 0, 2), lty=2)
> lines(x, dnorm(x, 0, 4), lty=3)
> legend(-9, 0.4, c(expression(sigma == 1), expression(sigma == 2), expression(sigma == 4.0)), lty=c(1,2,3))
```
Figure 4:
3.3.3. The Gamma Distribution

The *gamma* distribution with *shape parameter* $\alpha$ and *scale parameter* $\beta$ has density

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

where $\Gamma(\cdot)$ is the *gamma function* given by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

- It is required that both $\alpha > 0$ and $\beta > 0$.
- The mean and variance of the gamma are given by
  $$\mu = \alpha \beta \quad \text{and} \quad \sigma^2 = \alpha \beta^2.$$
- The gamma distribution is used for modeling random variables like survival times, waiting times, income, and rainfall amounts. Generally the distributions of positive, continuous random variables, especially those that are right-skewed, might be well-modeled by a gamma distribution.
- Different values of the shape parameter really do correspond to different shapes for the gamma the density (see Figure 5).
- Different scale parameters just affect the scale of measurement.
– As an example, suppose $X$ is the survival time of a light bulb, measured in hours, and that $X$ has a gamma distribution with $\alpha = 1$ and $\beta = 2000$. If we let $Y = 60X$, so that $Y$ is the survival time in seconds, then $Y$ has a gamma distribution with $\alpha = 1$ and $\beta = 60 \times 2000 = 12000$. 
Figure 5: Gamma densities with scale parameter $\beta = 1$ and shape parameters $\alpha = 0.75$, 1.0, and 2.0.
The following R functions are provided for the gamma distribution:

- **dgamma**: density function.
- **pgamma**: distribution function.
- **qgamma**: quantile function.
- **rgamma**: generate gamma random variables.

Figure 5 was produced by the following code:

```r
> x <- seq(0, 6, length=100)
> plot(x, dgamma(x, 0.75, 1), type="l", ylab="f(x)", lty=2)
> lines(x, dgamma(x, 1, 1), lty=1)
> lines(x, dgamma(x, 2, 1), lty=3)
> legend(4.0, 1.5, c(expression(alpha == 0.75),
                    expression(alpha == 1.00), expression(alpha == 2.0)),
        lty=c(2,1,3))
```

There are two important special cases of the gamma distribution: the exponential distribution and the chi-square distribution.
### 3.3.4. The Exponential Distribution

The *exponential* distribution is just a gamma distribution with $\alpha = 1$. Thus the exponential distribution with mean $\beta$ has density

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0.$$  

- The mean and variance of the exponential are given by
  $$\mu = \beta \quad \text{and} \quad \sigma^2 = \beta^2.$$  

- The exponential is often used to model waiting times and survival times. It has the *memoryless property*.

- The exponential distribution is often specified by giving its (hazard) rate, $\lambda = 1/\beta$ instead of $\beta$. Of course in this case the density is
  $$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$
3.3.5. The Chi-Square Distribution

The *chi-square* distribution with \( k > 0 \) degrees of freedom is just a gamma distribution with

\[
\alpha = \frac{k}{2} \quad \text{and} \quad \beta = 2.
\]

- The chi-square distribution is primarily important in statistical inference, as we shall see.

The following R functions are provided for the chi-square distribution:

- **dchisq**: density function.
- **pchisq**: distribution function.
- **qchisq**: quantile function.
- **rchisq**: generate chi-square random variables.
3.3.6. Student’s *t* Distribution

Student’s *t* distribution is another distribution that is primarily important in statistical inference.

- The *t* distribution with *k* degrees of freedom is best “operationally.” If
  - *Z* is a standard normal r.v.,
  - *X* is a chi-square r.v. with *k* d.f., and
  - *Z* and *X* are independent,

  then the random variable

  \[ T = \frac{Z}{\sqrt{X/k}} \]

  has a *t* distribution with *k* d.f.

- The *t* distribution looks like the standard normal but has heavier tails, i.e., the *t* distribution places more probability far from zero and less near zero than the standard normal.

- As the degrees of freedom grow larger, the *t* distribution gets closer and closer to the normal (see Figure 6).

- The *t* distribution with *k* d.f. has mean zero (if *k* ≥ 2) and variance

  \[ \sigma^2 = \frac{k}{k-2} \quad \text{if } k \geq 3. \]
Figure 6: Student’s $t$ distribution with 2 and 8 degrees of freedom compared to the standard normal.
The following R functions are provided for the $t$ distribution:

\textbf{dt}: density function.

\textbf{pt}: distribution function.

\textbf{qt}: quantile function.

\textbf{rt}: generate $t$ distributed random variables.

Figure 6 was produced by the following code:

\begin{verbatim}
> plot(x, dnorm(x), type="l", ylab="f(x)")
> lines(x, dt(x, 2), lty=2)
> lines(x, dt(x, 8), lty=3)
> legend(2, 0.4, c("d.f. = 2", "d.f. = 8", "normal"), lty=c(2,3,1))
\end{verbatim}
3.3.7. The $F$ Distribution

The $F$ distribution is yet another distribution that is primarily important in statistical inference.

- The $F$ distribution with $k_1$ numerator degrees of freedom and $k_2$ denominator degrees of freedom is best “operationally.” If
  
  - $X_1$ is a chi-square r.v. with $k_1$ d.f., and
  - $X_2$ is a chi-square r.v. with $k_2$ d.f., and
  - $X_1$ and $X_2$ are independent,

  then the random variable
  
  $$ F = \frac{X_1/k_1}{X_2/k_2} $$

  has an $F$ distribution with $k_1$ numerator and $k_2$ denominator d.f.

  The following R functions are provided for the $F$ distribution:

  - `df`: density function.
  - `pf`: distribution function.
  - `qf`: quantile function.
  - `rf`: generate $F$ distributed random variables.
3.3.8. The Beta Distribution

The Beta distribution with parameters $\alpha$ and $\beta$ has density

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1,$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

is the so-called beta function.

- The beta distribution is useful for modeling random proportions.
- Both $\alpha > 0$ and $\beta > 0$ must hold.
- A variety of shapes are possible by changing $\alpha$ and $\beta$.
- The beta distribution has mean and variance given by

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \sigma^2 = \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$
The following R functions are provided for the beta distribution:

- `dbeta`: density function.
- `pbeta`: distribution function.
- `qbeta`: quantile function.
- `rbeta`: generate $F$ distributed random variables.
3.3.9. Other Continuous Distributions

Some other common continuous distributions are the Cauchy, the logistic, the Weibull, and the log-normal. The R names for these are `cauchy`, `logis`, `weibull`, and `lnorm`, respectively, with the obvious four functions available for each, e.g., `dcauchy`, `pcauchy`, `qcauchy`, and `rcauchy`.

The Weibull and the log-normal are used for many of the same purposes as the Gamma distribution discussed earlier.