Topic 9

Examples of Mass Functions and Densities

For a given **state space**, S, we will describe several of the most frequently encountered parameterized families of both discrete and continuous random variables

$$X : \Omega \to S.$$

indexed by some parameter θ. We will add the subscript θ to the notation P_θ to indicate the parameter value used to compute probabilities. This section is meant to serve as an introduction to these families of random variables and not as a comprehensive development. The section should be considered as a reference to future topics that rely on this information.

We shall use the notation

$$f_X(x|\theta)$$

both for a family of mass functions for discrete random variables and the density functions for continuous random variables that depend on the parameter θ. After naming the family of random variables, we will use the expression $Family(\theta)$ as shorthand for this family followed by the R command and state space S. A table of R commands, parameters, means, and variances is given at the end of this situation.

9.1 Examples of Discrete Random Variables

Incorporating the notation introduced above, we write

$$f_X(x|\theta) = P_\theta\{X = x\}$$

for the mass function of the given family of discrete random variables.

1. (Bernoulli) $Ber(p)$, $S = \{0, 1\}$

$$f_X(x|p) = \left\{ \begin{array}{ll} 0 & \text{with probability } (1-p), \\ 1 & \text{with probability } p, \end{array} \right\} = p^x(1-p)^{1-x}.$$

   This is the simplest random variable, taking on only two values, namely, 0 and 1. Think of it as the outcome of a Bernoulli trial, i.e., a single toss of an unfair coin that turns up heads with probability p.

2. (binomial) $Bin(n, p)$ (R command `binom`) $S = \{0, 1, \ldots, n\}$

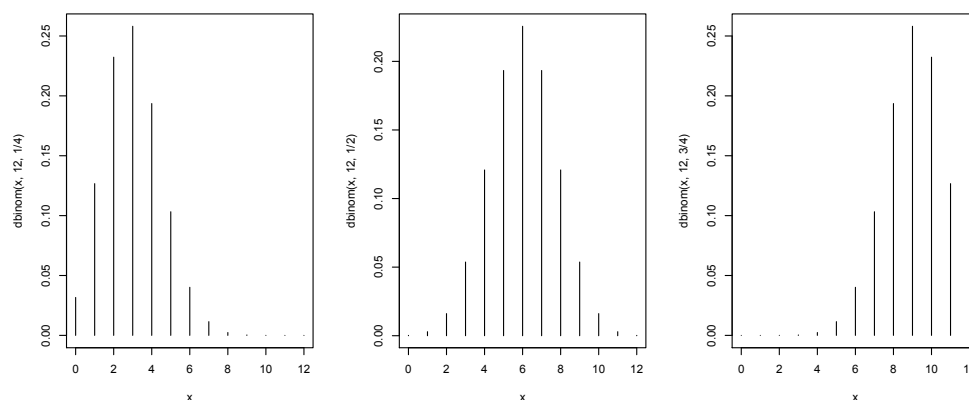$$f_X(x|p) = \binom{n}{x} p^x(1-p)^{n-x}.$$

Figure 9.1: Binomial mass function $f(x|p)$ for $n = 12$ and $p = 1/4, 1/2, 3/4$.

We gave a more extensive introduction to Bernoulli trials and the binomial distribution in the discussion on *The Expected Value*. Here we found that the binomial distribution arises from computing the probability of x successes in n Bernoulli trials. Considered in this way, the family $Ber(p)$ is also $Bin(1, p)$.

Notice that by its definition if X_i is $Bin(n_i, p), i = 1, 2$ and are independent, then $X_1 + X_2$ is $Bin(n_1 + n_2, p)$

3. (geometric) $Geo(p)$ (R command `geom`) $S = \mathbb{N} = \{0, 1, 2, \ldots\}$.

$$f_X(x|p) = p(1 - p)^x.$$

We previous described random variable as the number of failed Bernoulli trials before the first success. The name geometric random variable is also to the number of Bernoulli trials Y until the first success. Thus, $Y = X + 1$. As a consequence of these two choices for a geometric random variable, care should be taken to be certain which definition is under considertion.

Exercise 9.1. *Give the mass function for* Y.

4. (negative binomial) $Negbin(n, p)$ (R command `nbinom`) $S = \mathbb{N}$

$$f_X(x|p) = \binom{n + x - 1}{x} p^n (1 - p)^x.$$

This random variable is the number of failed Bernoulli trials before the n-th success. Thus, the family of geometric random variable $Geo(p)$ can also be denoted $Negbin(1, p)$. As we observe in our consideration
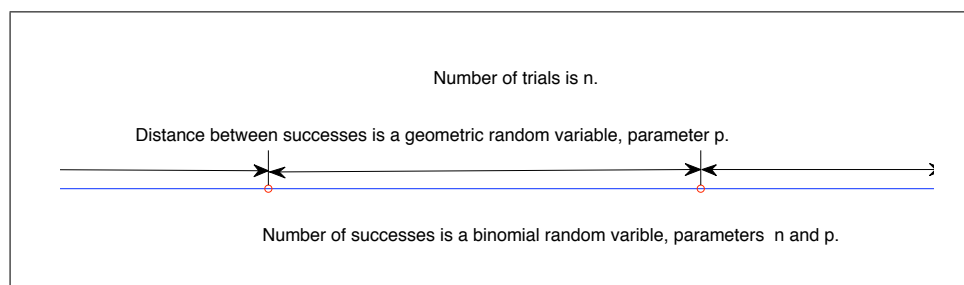

Figure 9.2: The relationship between the binomial and geometric random variable in Bernoulli trials.

138

of Bernoulli trials, we see that the number of failures between consecutive successes is a geometric random variable. In addition, the number of failures between any two pairs of successes (say, for example, the 2nd and 3rd success and the 6th and 7th success) are independent. In this way, we see that $Negbin(n, p)$ is the sum of n independent $Geo(p)$ random variables.

To find the mass function, note that in order for X to take on a given value x, then the n-th success must occur on the $n + x$-th trial. In other words, we must have $n - 1$ successes and x failures in first $n + x - 1$ Bernoulli trials followed by success on the last trial. The first $n + x - 1$ trials and the last trial are independent and so their probabilities multiply.

$$
\begin{aligned}
P_p\{X = x\} &= P_p\{n - 1 \text{ successes in } n + x - 1 \text{ trials, success in the } n - x\text{-th trial}\} \\
&= P_p\{n - 1 \text{ successes in } n + x - 1 \text{ trials}\}P_p\{\text{success in the } n - x\text{-th trial}\} \\
&= \binom{n + x - 1}{n - 1}p^{n-1}(1 - p)^x \cdot p = \binom{n + x - 1}{x}p^n(1 - p)^x
\end{aligned}
$$

The first factor is computed from the binomial distribution, the second from the Bernoulli distribution. Note the use of the identity

$$
\binom{m}{k} = \binom{m}{m - k}
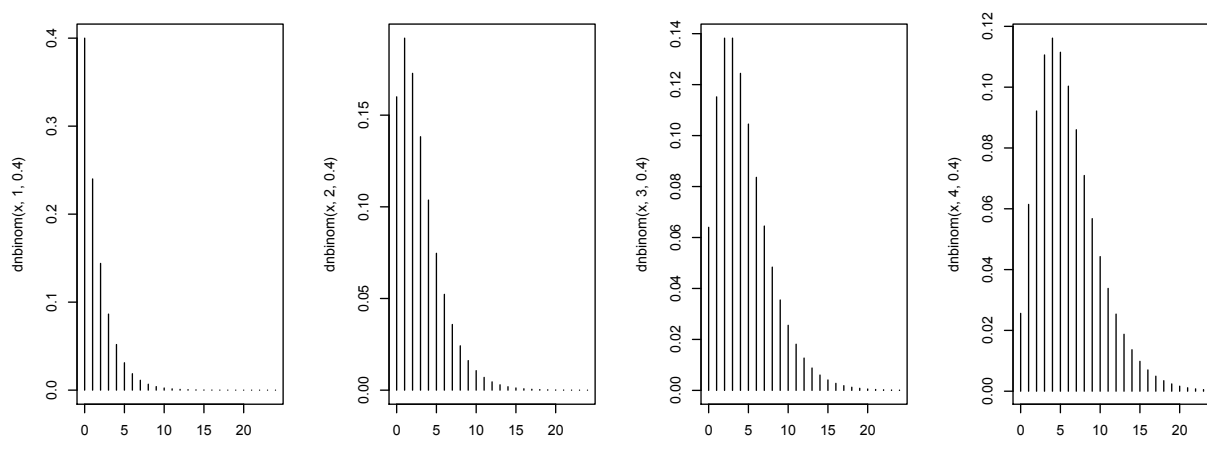$$

in giving the final formula.

Figure 9.3: Probability mass function for negative binomial random variables for $n = 1, 2, 3, 4$ and $p = 2/5$.

Exercise 9.2. *Use the fact that a negative binomial random variable $Negbin(r, p)$ is the sum of independent geometric random variable $Geo(p)$ to find its mean and variance. Use the fact that a geometric random variable has mean $(1 - p)/p$ and variance $(1 - p)/p^2$.*

5. (Poisson) $Pois(\lambda)$ (R command `pois`) $S = \mathbb{N}$,

$$
f_X(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}.
$$

The Poisson distribution approximates of the binomial distribution when n is large, p is small, but the product $\lambda = np$ is moderate in size. One example for this can be seen in bacterial colonies. Here, n is the number of bacteria and p is the probability of a mutation and λ, the mean number of mutations is moderate. A second is the

number of recombination events occurring during meiosis. In this circumstance, n is the number of nucleotides on a chromosome and p is the probability of a recombination event occurring at a particular nucleotide.

The approximation is based on the limit

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda} \tag{9.1}$$

We now compute binomial probabilities, replace p by λ/n and take a limit as $n \to \infty$. In this computation, we use the fact that for a fixed value of x,

$$\frac{(n)_x}{n^x} \to 1 \quad \text{and} \quad \left(1 - \frac{\lambda}{n}\right)^{-x} \to 1 \quad \text{as } n \to \infty$$

$$P\{X = 0\} = \binom{n}{0}p^0(1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}$$

$$P\{X = 1\} = \binom{n}{1}p^1(1-p)^{n-1} = n\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{n-1} \approx \lambda e^{-\lambda}$$

$$P\{X = 2\} = \binom{n}{2}p^2(1-p)^{n-2} = \frac{n(n-1)}{2}\left(\frac{\lambda}{n}\right)^2\left(1 - \frac{\lambda}{n}\right)^{n-2} = \frac{n(n-1)}{n^2}\frac{\lambda^2}{2}\left(1 - \frac{\lambda}{n}\right)^{n-2} \approx \frac{\lambda^2}{2}e^{-\lambda}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$P\{X = x\} = \binom{n}{x}p^x(1-p)^{n-x} = \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{(n)_x}{n^x}\frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^{n-x} \approx \frac{\lambda^x}{x!}e^{-\lambda}.$$

The Taylor series for the exponential function

$$\exp \lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}.$$

shows that

$$\sum_{x=0}^{\infty} f_X(x) = 1.$$

Exercise 9.3. *Take logarithms and use l'Hôpital's rule to establish the limit (9.1) above.*

Exercise 9.4. *We saw that the sum of independent binomial random variables with a common value for p, the success probability, is itself a binomial random variable. Show that the sum of independent Poisson random variables is itself a Poisson random variable. In particular, if X_i are $Pois(\lambda_i), i = 1, 2$, then $X_1 + X_2$ is $Pois(\lambda_1 + \lambda_2)$.*

6. (uniform) $U(a, b)$ (R command `sample`) $S = \{a, a + 1, \ldots, b\}$,

$$f_X(x|a, b) = \frac{1}{b - a + 1}.$$

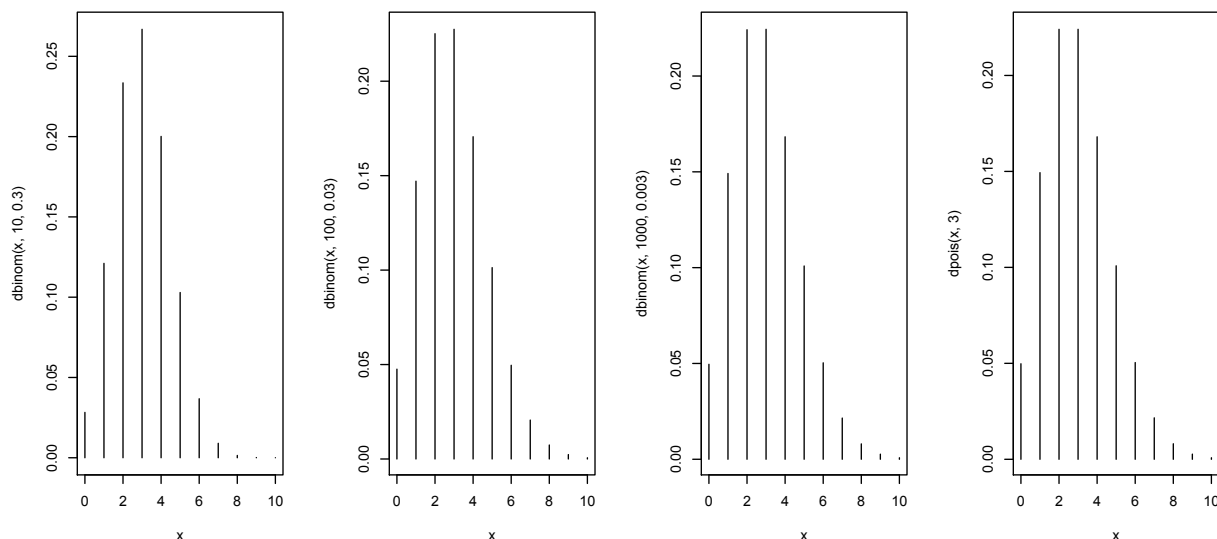Thus each value in the designated range has the same probability.

Figure 9.4: Probability mass function for binomial random variables for (a) $n = 10, p = 0.3$, (b) $n = 100, p = 0.03$, (c) $n = 1000, p = 0.003$ and for (d) the Poisson random varialble with $\lambda = np = 3$. This displays how the Poisson random variable approximates the binomial random variable with n large, p small, and their product $\lambda = np$ moderate.

7. (hypergeometric) $Hyper(m, n, k)$ (R command hyper). The hypergeometric distribution will be used in computing probabilities under circumstances that are associated with sampling without replacement. We will use the analogy of an urn containing balls having one of two possible colors.

Begin with an urn holding m white balls and n black balls. Remove k and let the random variable X denote the number of white balls. The value of X has several restrictions. X cannot be greater than either the number of white balls, m, or the number chosen k. In addition, if $k > n$, then we must consider the possibility that all of the black balls were chosen. If $X = x$, then the number of black balls, $k - x$, cannot be greater than the number of black balls, n, and thus, $k - x \leq n$ or $x \geq k - n$.

If we are considering equally likely outcomes, then we first compute the total number of possible outcomes, $\#(\Omega)$, namely, the number of ways to choose k balls out of an urn containing $m + n$ balls. This is the number of combinations

$$\binom{m+n}{k}.$$

This will be the denominator for the probability. For the numerator of $P\{X = x\}$, we consider the outcomes that result in x white balls from the total number m in the urn. We must also choose $k - x$ black balls from the total number n in the urn. By the multiplication property, the number of ways $\#(A_x)$ to accomplish this is product of the number of outcomes for these two combinations,

$$\binom{m}{x}\binom{n}{k-x}.$$

The mass function for X is the ratio of these two numbers.

$$f_X(x|m, n, k) = \frac{\#(A_x)}{\#(\Omega)} = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}, \quad x = \max\{0, k-n\}, \ldots, \min\{m, k\}.$$

141

Exercise 9.5. *Show that we can rewrite this probability as*

$$f_X(x|m,n,k) = \frac{k!}{x!(k-x)!}\frac{(m)_x(n)_{k-x}}{(m+n)_k} = \binom{k}{x}\frac{(m)_x(n)_{k-x}}{(m+n)_k}. \tag{9.2}$$

This gives probabilities using sampling **without replacement**. If we were to choose the balls one-by-one returning the balls to the urn after each choice, then we would be sampling **with replacement**. This returns us to the case of k Bernoulli trials with success parameter $p = m/(m+n)$, the probability for choosing a white ball. In the case the mass function for Y, the number of white balls, is

$$f_Y(x|m,n,k) = \binom{k}{x}p^x(1-p)^{k-x} = \binom{k}{x}\left(\frac{m}{m+n}\right)^x\left(\frac{n}{m+n}\right)^{k-x} = \binom{k}{x}\frac{m^x n^{k-x}}{(m+n)^k}. \tag{9.3}$$

Note that the difference in the formulas between sampling with replacement in (9.3) and without replacement in (9.2) is that the powers are replaced by the falling function, e.g., m^x is replaced by $(m)_x$.

Let X_i be a Bernoulli random variable indicating whether or not the color of the i-th ball is white. Thus, its mean

$$EX_i = \frac{m}{m+n}.$$

The random variable for the total number of white balls $X = X_1 + X_2 + \cdots + X_k$ and thus its mean

$$EX = EX_1 + EX_2 + \cdots + EX_k = k\frac{m}{m+n}.$$

Because the selection of white for one of the marbles decreases the chance for black for another selection, the trials are not independent. One way to see this is by noting the variance (not derived here) of the sum $X = X_1 + X_2 + \cdots + X_k$

$$\text{Var}(X) = k\frac{m}{m+n}\frac{n}{m+n}\cdot\frac{m+n-k}{m+n-1}$$

is not the sum of the variances.

If we write $N = m + n$ for the total number of balls in the urn and $p = m/(m+n)$ as above, then

$$\text{Var}(X) = kp(1-p)\frac{N-k}{N-1}$$

Thus the variance of the hypergeometric random variable is reduced by a factor of $(N-k)/(N-1)$ from the case of the corresponding binomial random variable. In the cases for which k is much smaller than N, then sampling with and without replacement are nearly the same process - any given ball is unlikely to be chosen more than once under sampling with replacement. We see this situation, for example, in a opinion poll with k at 1 or 2 thousand and n, the population of a country, typically many millions.

On the the other hand, if k is a significant fraction of N, then the variance is significantly reduced under sampling without replacement. We are much less uncertain about the fraction of white and black balls. In the extreme case of $k = N$, we have chosen every ball and know that $X = m$ with probability 1. In the case, the variance formula gives $\text{Var}(X) = 0$, as expected.

Exercise 9.6. *Draw two balls without replacement from the urn described above. Let X_1, X_2 be the Bernoulli random indicating whether or not the ball is white. Find $Cov(X_1, X_2)$.*

Exercise 9.7. *Check that $\sum_{x \in S} f_X(x|\theta) = 1$ in the examples above.*

9.2 Examples of Continuous Random Variables

For continuous random variables, we have for the density

$$f_X(x|\theta) \approx \frac{P_\theta\{x < X \le x + \Delta x\}}{\Delta x}.$$

1. (uniform) $U(a,b)$ (R command `unif`) on $S = [a,b]$,
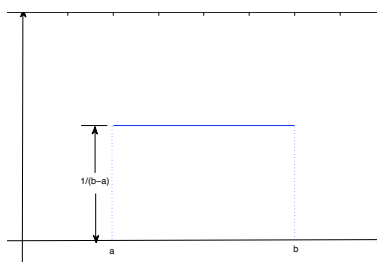
$$f_X(x|a,b) = \frac{1}{b-a}.$$

Figure 9.5: Uniform density

Independent $U(0,1)$ are the most common choice for generating random numbers. Use the R command `runif(n)` to simulate n independent random numbers.

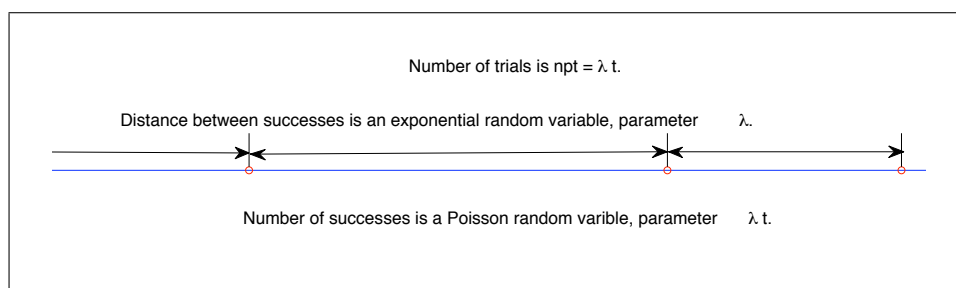**Exercise 9.8.** *Find the mean and the variance of a $U(a,b)$ random variable.*

Figure 9.6: The relationship between the Poission and exponential random variable in Bernoulli trials with large n, small p and moderate size product $\lambda = np$. Notice the analogies from Figure 9.2. Imagine a bacterial colony with individual bacterium produced at a constant rate n per unit time. Then, the times between mutations can be approximated by independent exponential random variables and the number of mutations is approximately a Poisson random variable.

2. (exponential) $Exp(\lambda)$ (R command `exp`) on $S = [0,\infty)$,

$$f_X(x|\lambda) = \lambda e^{-\lambda x}.$$

To see how an exponential random variable arises, consider Bernoulli trials arriving at a rate of n trials per time unit and take the approximation seen in the Poisson random variable. Again, the probability of success p is small, ns the number of trials up to a given time s is large, and $\lambda = np$. Let T be the time of the first success. This random time exceeds a given time s if we begin with ns consecutive failures. Thus, the survival function

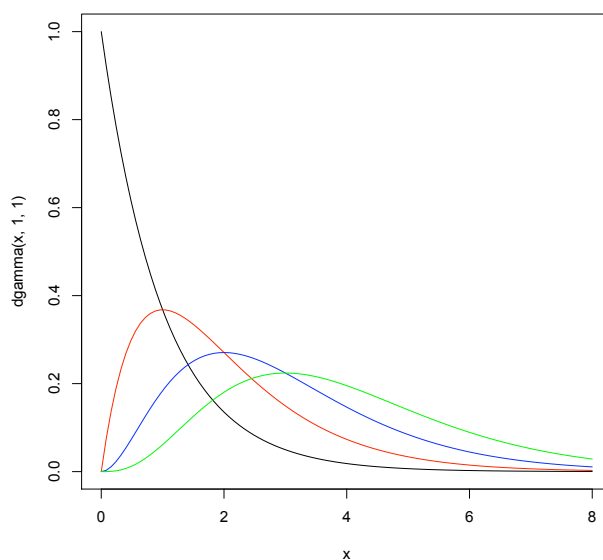$$\bar{F}_T(s) = P\{T > s\} = (1-p)^{ns} = \left(1 - \frac{\lambda}{n}\right)^{ns} \approx e^{-\lambda s}.$$

Figure 9.7: Density for a gamma random variable. Here, $\beta = 1$, $\alpha = 1$ (black), 2 (red), 3 (blue) and 4 (green)

The cumulative distribution function

$$F_T(s) = P\{T \le s\} = 1 - P\{T > s\} \approx 1 - e^{-\lambda s}.$$

The density above can be found by taking a derivative of $F_T(s)$.

3. (gamma) $\Gamma(\alpha, \beta)$ (R command gamma) on $S = [0, \infty)$,

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

Observe that $Exp(\lambda)$ is $\Gamma(1, \lambda)$. A $\Gamma(n, \lambda)$ can be seen as an approximation to the negative binomial random variable using the ideas that leads from the geometric random variable to the exponential. Alternatively, for a natural number n, $\Gamma(n, \lambda)$ is the sum of n independent $Exp(\lambda)$ random variables. This special case of the gamma distribution is sometimes called the **Erlang distribution** and was originally used in models for telephone traffic.

The **gamma function**

$$\Gamma(s) = \int_0^\infty x^s e^{-x} \frac{dx}{x}$$

This is computed in R using gamma(s).

For the graphs of the densities in Figure 9.7,

```
> curve(dgamma(x,1,1),0,8)
> curve(dgamma(x,2,1),0,8,add=TRUE,col="red")
> curve(dgamma(x,3,1),0,8,add=TRUE,col="blue")
> curve(dgamma(x,4,1),0,8,add=TRUE,col="green")
```

Exercise 9.9. *Use integration by parts to show that*

$$\Gamma(t+1) = t\Gamma(t). \tag{9.4}$$

If n is a non-negative integer, show that $\Gamma(n) = (n-1)!$.
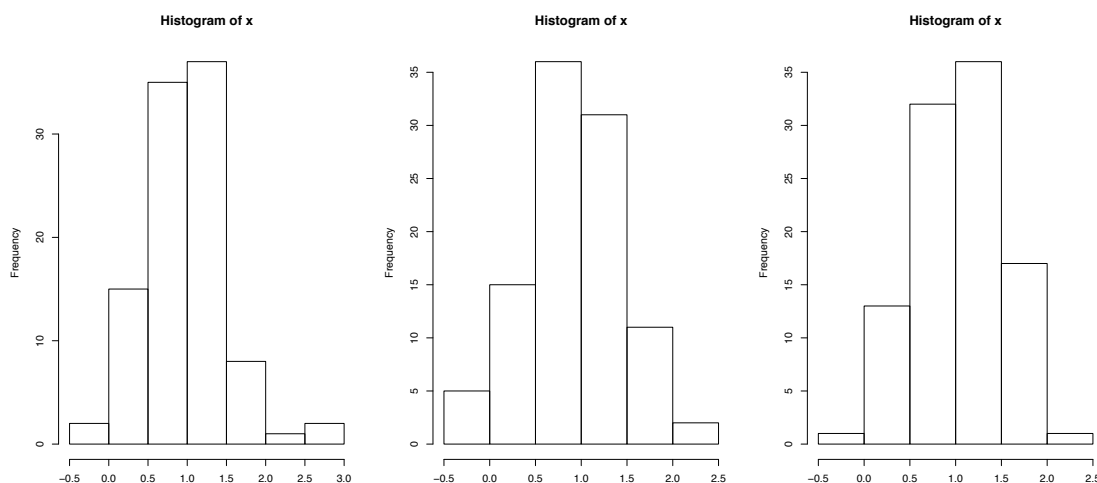
144

Figure 9.8: Histrogram of three simulations of 200 normal random variables, mean 1, standard deviation 1/2

4. (beta) $Beta(\alpha, \beta)$ (R command `beta`) on $S = [0, 1]$,

$$f_X(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

Beta random variables appear in a variety of circumstances. One common example is the **order statistics**. Beginning with n observations, $X_1, X_2, \cdots, X_n$, of independent uniform random variables on the interval $[0, 1]$ and rank them

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

from smallest to largest. Then, the k-th order statistic $X_{(k)}$ is $Beta(k, n - k + 1)$.

Exercise 9.10. *Use the identity (9.4) for the gamma function to find the mean and variance of the beta distribution.*

5. (normal) $N(\mu, \sigma)$ (R command `norm`) on $S = \mathbb{R}$,

$$f_X(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

Thus, a standard normal random variable is $N(0, 1)$. Other normal random variables are linear transformations of Z, the standard normal. In particular, $X = \sigma Z + \mu$ has a $N(\mu, \sigma)$ distribution. To simulate 200 normal random variables with mean 1 and standard deviation 1/2, use the R command `x<-rnorm(200,1,0.5)`. Histograms of three simulations are given in the Figure 9.8.

We often move from a random variable X to a function g of the random variable. Thus, $Y = g(X)$. The following exercise give the density f_Y in terms of the density f_X for X in the case that g is a monotone function.

Exercise 9.11. *Show that for g differentiable and monotone then g has a differentiable inverse g^{-1} and the density*

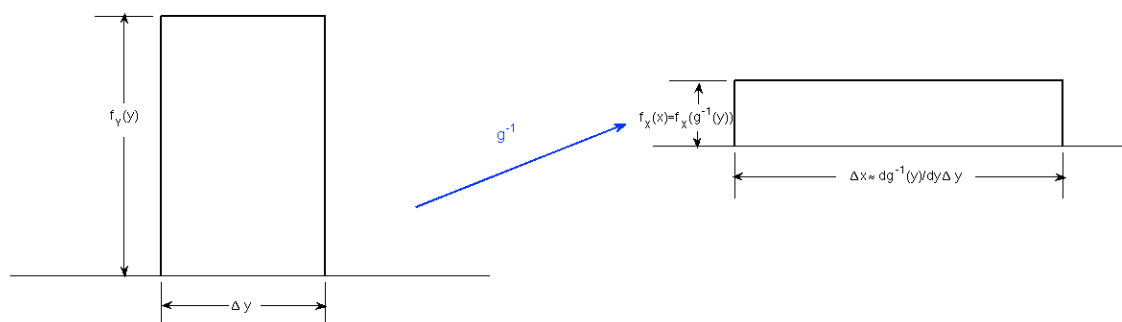$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|. \tag{9.5}$$

Figure 9.9: Finding the density of $Y = g(X)$ from the density of X. The areas of the two rectangles should be the same. Consequently, $f_Y(y)\Delta y \approx f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)\Delta y$.

We can find this density geometrically by noting that for g increasing,

Y is between y and $y + \Delta y$ if and only if X is between $g^{-1}(y)$ and $g^{-1}(y + \Delta y) \approx g^{-1}(y) + \dfrac{d}{dy}g^{-1}(y)\Delta y$.

Thus,

$$f_Y(y)\Delta y = P\{y < Y \le y + \Delta y\} \approx P\{g^{-1}(y) < X \le g^{-1}(y) + \frac{d}{dy}g^{-1}(y)\Delta y\}$$

$$\approx f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)\Delta y.$$

Dropping the factors Δy gives (9.5). (See Figure 9.9.)

6. (log-normal) $\ln N(\mu, \sigma)$ (R command `lnorm`) on $S = (0, \infty)$. A **log-normal** random variable is the exponential of a normal random variable. Thus, the logarithm of a log-normal random variable is normal. The density of this family is
$$f_X(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right).$$

Exercise 9.12. *Use the exercise above to find the density of a log-normal random variable.*

As we stated previously, the normal family of random variables will be the most important collection of distributions we will use. For example, the final three examples, each of them derived from normal random variables, will be seen as densities of statistics used in hypothesis testing. Even though their densities are given explicitly, in practice, these formulas are rarely used directly. Probabilities are generally computed using software.

7. (chi-squared) χ_a^2 (R command `chisq`) on $S = [0, \infty)$
$$f_X(x|a) = \frac{x^{a/2-1}}{2^{a/2}\Gamma(a/2)}e^{-x/2}.$$

The value a is called the number of **degrees of freedom**. For a a positive integer, let $Z_1, Z_2, \ldots, Z_a$ be independent standard normal random variables. Then,
$$Z_1^2 + Z_2^2 + \cdots + Z_a^2$$

has a χ_a^2 distribution.

Exercise 9.13. *Modify the solution to the exercise above to find the density of a χ_1^2 random variable. (Hint: For $Y = X^2$, $P\{Y \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\}$.)*

8. (Student's t) $t_a(\mu, \sigma^2)$ (R command t) on $S = \mathbb{R}$,

$$f_X(x|a) = \frac{\Gamma((a+1)/2)}{\sqrt{\alpha\pi}\Gamma(\alpha/2)\sigma}\left(1 + \frac{(x-\mu)^2}{a\sigma^2}\right)^{-(a+1)/2}.$$

The value a is also called the number of **degrees of freedom**. If $\bar{Z}$ is the sample mean of n standard normal random variables and

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

is the sample variance, then

$$T = \frac{\bar{Z}}{S/\sqrt{n}}.$$

has a $t_{n-1}(0, 1)$ distribution.

9. (Fisher's F) $F_{q,a}$ (R command f) on $S = [0, \infty)$,

$$f_X(x|q, a) = \frac{\Gamma((q+a)/2)q^{q/2}a^{a/2}}{\Gamma(q/2)\Gamma(a/2)}x^{q/2-1}(a + qx)^{-(q+a)/2}.$$

The F distribution will make an appearance when we see the analysis of variance test. It arises as the ratio of independent chi-square random variables. The chi-square random variable in the numerator has a degrees of freedom; the chi-square random variable in the denominator has q degrees of freedom

9.3 R **Commands**

R can compute a variety of values for many families of distributions.

- `dfamily(x, parameters)` is the mass function (for discrete random variables) or probability density (for continuous random variables) of `family` evaluated at x.

- `qfamily(p, parameters)` returns x satisfying $P\{X \leq x\} = p$, the p-**th quantile** where X has the given distribution,

- `pfamily(x, parameters)` returns $P\{X \leq x\}$ where X has the given distribution.

- `rfamily(n, parameters)` generates n random variables having the given distribution.

9.4 **Summary of Properties of Random Variables**

9.4.1 **Discrete Random Variables**

random variable	R	parameters	mean	variance	generating function
Bernoulli	*	p	p	$p(1-p)$	$(1-p) + pz$
binomial	binom	n, p	np	$np(1-p)$	$((1-p) + pz)^n$
geometric	geom	p	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)z}$
hypergeometric	hyper	m, n, k	$\frac{km}{m+n}$	$k\frac{m}{m+n} \cdot \frac{n}{m+n} \cdot \frac{m+n-k}{m+n-1}$	
negative binomial	nbinom	a, p	$a\frac{1-p}{p}$	$a\frac{1-p}{p^2}$	$\left(\frac{p}{1-(1-p)z}\right)^a$
Poisson	pois	λ	λ	λ	$\exp(-\lambda(1-z))$
uniform	sample	a, b	$\frac{b-a+1}{2}$	$\frac{(b-a+1)^2-1}{12}$	$\frac{z^a}{b-a+1}\frac{1-z^{b-a+1}}{1-z}$

147

*For a Bernoulli random variable, use the binomial commands with `n=1` trial.

9.4.2   Continuous Random Variables

random variable	R	parameters	mean	variance	characteristic function
beta	`beta`	α, β	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$F_{1,1}(a, b; \frac{i\theta}{2\pi})$
chi-squared	`chisq`	a	a	$2a$	$\frac{1}{(1-2i\theta)^{a/2}}$
exponential	`exp`	λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{i\lambda}{\theta+i\lambda}$
log-normal	`lnorm`	μ, σ	$\exp(\mu + \sigma^2/2)$	$(e^{\sigma^2} - 1)\exp(2\mu + \sigma^2)$	
F	`f`	q, a	$\frac{a}{a-2}, a > 2$	$2a^2\frac{q+a-2}{q(a-4)(a-2)^2}$	
gamma	`gamma`	α, β	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(\frac{i\beta}{\theta+i\beta}\right)^\alpha$
normal	`norm`	μ, σ^2	μ	σ^2	$\exp(i\mu\theta - \frac{1}{2}\sigma^2\theta^2)$
t	`t`	a, μ, σ^2	$\mu, a > 1$	$\sigma^2\frac{a}{a-2}, a > 1$	
uniform	`unif`	a, b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$-i\frac{\exp(i\theta b)-\exp(i\theta a)}{\theta(b-a)}$

Example 9.14. *We give several short examples that use the* R *commands introduced above.*

- *To find the values of the mass function $f_X(x) = P\{X = x\}$ for a binomial random variable 4 trials with probability of success $p = 0.7$*

```
> x<-c(0:4)
> binomprob<-dbinom(x,4,0.7)
> data.frame(x,binomprob)
  x binomprob
1 0    0.0081
2 1    0.0756
3 2    0.2646
4 3    0.4116
5 4    0.2401
```

- *To find the probability $P\{X \leq 3\}$ for X a geometric random variable with probability of success $p = 0.3$ enter* `pgeom(3,0.3)`. R *returns the answer 0.7599.*

- *To find the deciles of a gamma random variable with $\alpha = 4$ and $\beta = 5$*

```
> decile<-seq(0,0.9,0.1)
> value<-qgamma(decile,4,5)
> data.frame(decile,value)
   decile     value
1     0.0 0.0000000
2     0.1 0.3489539
3     0.2 0.4593574
4     0.3 0.5527422
5     0.4 0.6422646
6     0.5 0.7344121
7     0.6 0.8350525
8     0.7 0.9524458
9     0.8 1.1030091
10    0.9 1.3361566
```

- *To give independent observations uniformly on a set S, use the* `sample` *command using* `replace=TRUE`. *Here is an example usingt 50 repeated rolls of a die*

```
> S<-c(1:6)
> x<-sample(S,50,replace=TRUE)
> x
 [1] 3 3 4 4 1 3 6 4 3 5 5 1 3 4 6 3 5 1 5 2 6 3 4 4 1 3 1 6 5 4 2 1
[33] 2 3 4 1 2 1 1 6 5 1 2 3 4 5 1 3 6 5
```

- *The command* `rnorm(200,1,0.5)` *was used to create the histograms in Figure 9.8.*

- *Use the* `curve` *command to plot density and distribution functions. Thus was accomplished in Figure 9.7 using* `dgamma` *for the density of a gamma random variable. For cumulative distribution functions use* `pdist` *and substiture for* `diet` *the appropriate command from the table above.*

9.5   Answers to Selected Exercises

9.1. For $y = 1, 2, \ldots$,

$$f_Y(y) = P\{Y = y\} = P\{X + 1 = y\} = P\{X = y - 1\} = p(1-p)^{y-1}.$$

9.2. Write X a $Negbin(n, p)$ random variable as $X = Y_1 + \cdots + Y_n$ where the Y_i are independent random variable. Then,

$$EX = EY_1 + \cdots + EY_n = \frac{1-p}{p} + \cdots + \frac{1-p}{p} = \frac{n(1-p)}{p}$$

and because the Y_i are independent

$$\text{Var}(X) = \text{Var}(Y_1) + \cdots + \text{Var}(Y_n) = \frac{1-p}{p^2} + \cdots + \frac{1-p}{p^2} = \frac{n(1-p)}{p^2}$$

9.3. By taking the logarithm, the limit above is equivalent to

$$\lim_{n \to \infty} n \ln\left(1 - \frac{\lambda}{n}\right) = -\lambda.$$

Now change variables letting $\epsilon = 1/n$, then the limit becomes

$$\lim_{\epsilon \to 0} \frac{\ln(1 - \epsilon\lambda)}{\epsilon} = -\lambda.$$

The limit has the indeterminant form 0/0. Thus, by l'Hôpital's rule, we can take the derivative of the numerator and denominator to obtain the equivalent problem

$$\lim_{\epsilon \to 0} \frac{-\lambda}{1 - \epsilon\lambda} = -\lambda.$$

9.4. We have mass functions

$$f_{X_1}(x_1) = \frac{\lambda_1^{x_1}}{x_1!}e^{-\lambda_1} \quad f_{X_2}(x_2) = \frac{\lambda_2^{x_2}}{x_2!}e^{-\lambda_2}$$

149

Thus,

$$
f_{X_1+X_2}(x) = P\{X_1 + X_2 = x\} = \sum_{x_1=0}^{x} P\{X_1 = x_1, X_2 = x - x_1\} = \sum_{x_1=0}^{x} P\{X_1 = x_1\}P\{X_2 = x - x_1\}
$$

$$
= \sum_{x_1=0}^{x} \frac{\lambda_1^{x_1}}{x_1!}e^{-\lambda_1}\frac{\lambda_2^{x-x_1}}{(x-x_1)!}e^{-\lambda_2} = \left(\sum_{x_1=0}^{x} \frac{1}{x_1!(x-x_1)!}\lambda_1^{x_1}\lambda_2^{x-x_1}\right)e^{-(\lambda_1+\lambda_2)}
$$

$$
= \left(\sum_{x_1=0}^{x} \frac{x!}{x_1!(x-x_1)!}\lambda_1^{x_1}\lambda_2^{x-x_1}\right)\frac{1}{x!}e^{-(\lambda_1+\lambda_2)} = \frac{(\lambda_1+\lambda_2)^x}{x!}e^{-(\lambda_1+\lambda_2)}.
$$

This is the probability mass function for a $Pois(\lambda_1+\lambda_2)$ random variable. The last equality uses the binomial theorem.

9.5. Using the definition of the choose function

$$
f_X(x|m,n,k) = \frac{\binom{b}{x}\binom{n}{k-x}}{\binom{m+n}{k}} = \frac{\frac{(m)_x}{x!}\frac{(n)_{k-x}}{(k-x)!}}{\frac{(m+n)_k}{k!}} = \frac{k!}{x!(k-x)!}\frac{(m)_x(n)_{k-x}}{(m+n)_k} = \binom{k}{x}\frac{(m)_x(n)_{k-x}}{(m+n)_k}.
$$

9.6. $\mathrm{Cov}(X_1, X_2) = EX_1X_2 - EX_1EX_2$. Now,

$$
EX_1 = EX_2 = \frac{m}{m+n} = p
$$

and

$$
EX_1X_2 = P\{X_1X_2 = 1\} = P\{X_2 = 1, X_1 = 1\} = P\{X_2 = 1|X_1 = 1\}P\{X_1 = 1\} = \frac{m-1}{m+n-1}\cdot\frac{m}{m+n}.
$$

Thus,

$$
\mathrm{Cov}(X_1, X_2) = \frac{m-1}{m+n-1}\cdot\frac{m}{m+n} - \left(\frac{m}{m+n}\right)^2 = \frac{m}{m+n}\left(\frac{m-1}{m+n-1} - \frac{m}{m+n}\right)
$$

$$
= \frac{m}{m+n}\left(\frac{(m+n)(m-1) - m(m+n-1)}{(m+n)(m+n-1)}\right) = \frac{m}{m+n}\left(\frac{-n}{(m+n)(m+n-1)}\right) = -\frac{np}{N(N-1)}
$$

where, as above, $N = m + n$.

9.8. For the mean

$$
E_{(a,b)}X = \int_a^b xf_X(x|a,b)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{2(b-a)}x^2\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2},
$$

the average of endpoints a and b. For the variance, we first find the second moment

$$
E_{(a,b)}X^2 = \int_a^b x^2f_X(x|a,b)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{3(b-a)}x^3\Big|_a^b = \frac{b^3-a^3}{3(b-a)}\frac{(b-a)(a^2+ab+b^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3}.
$$

Thus,

$$
\mathrm{Var}_{(a,b)}(X) = \frac{a^2+ab+b^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{4b^2+4ab+4b^2}{12} - \frac{3a^2+6ab+3a^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(a-b)^2}{12}.
$$

9.9. Using integration by parts

$$\begin{aligned}u(x) &= x^t & v(x) &= -e^{-x}\\ u'(z) &= tx^{t-1} & v'(x) &= e^{-x}\end{aligned}$$

$$\Gamma(t+1) = \int_0^\infty x^t e^{-x}\, dx = -x^t e^{-x}\Big|_0^\infty + t\int_0^\infty x^{t-1}e^{-x}\, dx = t\Gamma(t).$$

The first term is 0 because $x^t e^{-x} \to 0$ as $x \to \infty$.

For the case $n = 1$, $\Gamma(1) = \int_0^\infty e^{-s}\, ds = 1 = (1-1)!$. Now assume that $\Gamma(n) = (n-1)!$ We have just shown this identity in the case $n = 1$. Now,

$$\Gamma(n+1) = n\Gamma(n) = n \cdot (n-1)! = n!.$$

Thus, by induction we have the formula for all integer values.

9.10. In order to be a probability density, we have that

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx.$$

We use this identity and (9.4) to compute the first two moments

$$\begin{aligned}E_{(\alpha,\beta)}X &= \int_0^1 x f_X(x|\alpha,\beta)\, dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^\alpha(1-x)^{\beta-1}\, dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\\ &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)} = \frac{\Gamma(\alpha+\beta)\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)} = \frac{\alpha}{\alpha+\beta}.\end{aligned}$$

and

$$\begin{aligned}E_{(\alpha,\beta)}X^2 &= \int_0^1 x^2 f_X(x|\alpha,\beta)\, dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha+1}(1-x)^{\beta-1}\, dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}\\ &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha+\beta+2)\Gamma(\alpha)} = \frac{\Gamma(\alpha+\beta)(\alpha+1)\alpha\Gamma(\alpha)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)} = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.\end{aligned}$$

Thus,

$$\begin{aligned}\text{Var}_{(\alpha,\beta)}(X) &= E_{(\alpha,\beta)}X^2 - (E_{(\alpha,\beta)}X)^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2\\ &= \frac{(\alpha+1)\alpha(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}\end{aligned}$$

9.11. We first consider the case of g increasing on the range of the random variable X. In this case, g^{-1} is also an increasing function.

To compute the cumulative distribution of $Y = g(X)$ in terms of the cumulative distribution of X, note that

$$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\} = P\{X \le g^{-1}(y)\} = F_X(g^{-1}(y)).$$

Now use the chain rule to compute the density of Y

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}F_X(g^{-1}(y)) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y).$$

For g decreasing on the range of X,

$$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\} = P\{X \ge g^{-1}(y)\} = 1 - F_X(g^{-1}(y)),$$

151

and the density

$$f_Y(y) = F_Y'(y) = -\frac{d}{dy}F_X(g^{-1}(y)) = -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y).$$

For g decreasing, we also have g^{-1} decreasing and consequently the density of Y is indeed positive,

We can combine these two cases to obtain

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|.$$

9.12. Let X be a normal random variable, then $Y = \exp X$ is log-normal. Thus $y = g(x) = e^x$, $g^{-1}(y) = \ln y$, and $\frac{d}{dy}g^{-1}(y) = \frac{1}{y}$. Note that y must be positive. Thus,

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right)\frac{1}{y}.$$

9.13. Let X be a standard normal random variable, then $Y = X^2$ is χ_1^2. From the hint, the distribution function of Y,

$$F_Y(y) = P\{Y \le y\} = P\{-\sqrt{y} \le X \le \sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Now take a derivative with respect to y.

$$f_Y(y) = P\{Y \le y\} = f_X(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right)$$

$$= (f_X(\sqrt{y}) + f_X(-\sqrt{y}))\frac{1}{2\sqrt{y}} = \left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y}{2}\right) + \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y}{2}\right)\right)\frac{1}{2\sqrt{y}}$$

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

Finally, $\Gamma(1/2) = \sqrt{\pi}$.