## 3.6 Joint Distributions

Properties of the *joint* (*bivariate*) *discrete probability mass function* pmf $f(x, y) = P(X = x, Y = y)$ for random variables $X$ and $Y$ with ranges $R_X$ and $R_Y$ where $R = \{(x, y) | x \in R_X, y \in R_Y\}$, are:

- $0 \leq f(x, y) \leq 1$, for all $x \in R_X, y \in R_Y$,
- $\sum\sum_{(x,y)\in R} f(x, y) = 1$,
- if $S \subset R$, $P[(X, Y) \in S] = \sum\sum_{(x,y)\in S} f(x, y)$,

with *marginal* pmfs of $X$ and of $Y$,

$$f_X(x) = P(X = x) = \sum_{y \in R_Y} f(x, y), \quad f_Y(y) = P(Y = y) = \sum_{x \in R_X} f(x, y).$$

Properties of the *joint* (*bivariate*) *continuous probability density function* pdf $f(x, y)$ for continuous random variables $X$ and $Y$, are:

- $f(x, y) \geq 0$, $-\infty < x < \infty, -\infty < y < \infty$,
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1$,
- if $S$ is a subset of two-dimensional plane, $P[(X, Y) \in S] = \int\int_S f(x, y)\, dy\, dx$,

with *marginal* pdfs of $X$ and of $Y$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Random variables (discrete or continuous) $X$ and $Y$ are *independent* if and only if

$$f(x, y) = f_X(x) \cdot f_Y(y).$$

A set of $n$ random variables are *mutually independent* if and only if

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_3) \cdots f_{X_n}(x_n).$$

**Exercise 3.6 (Joint Distributions)**

1. *Discrete joint (bivariate) pmf: marbles drawn from an urn.* Marbles chosen at random with*out* replacement from an urn consist of 8 blue and 6 black marbles. Blue counts for 0 points and black counts for 1 point. Let $X$ denote number of points from first marble chosen and $Y$ denote number of points from second marble chosen.

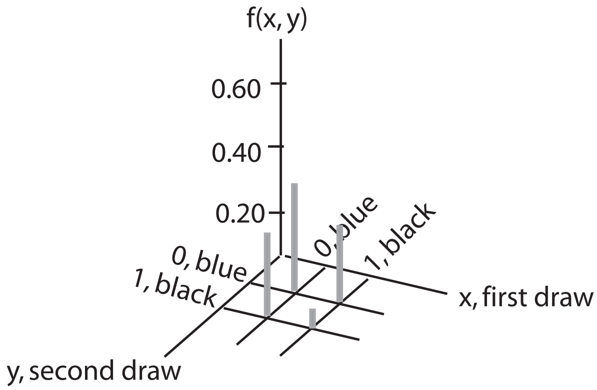



Figure 3.14: Discrete bivariate function: marbles

(a) Chance of choosing two blue marbles is $f(x,y) = f(0,0) =$
(i) $\frac{\mathbf{8 \cdot 7}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{28}}{\mathbf{91}}$ (ii) $\frac{\mathbf{8 \cdot 6}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{24}}{\mathbf{91}}$ (iii) $\frac{\mathbf{6 \cdot 8}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{24}}{\mathbf{91}}$ (iv) $\frac{\mathbf{6 \cdot 5}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{15}}{\mathbf{91}}$.

(b) Chance of a blue marble then black marble is $f(x,y) = f(0,1) =$
(i) $\frac{\mathbf{8 \cdot 7}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{28}}{\mathbf{91}}$ (ii) $\frac{\mathbf{8 \cdot 6}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{24}}{\mathbf{91}}$ (iii) $\frac{\mathbf{6 \cdot 8}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{24}}{\mathbf{91}}$ (iv) $\frac{\mathbf{6 \cdot 5}}{\mathbf{14 \cdot 13}} = \frac{\mathbf{15}}{\mathbf{91}}$.

(c) The joint density is

| first drawn, $x$ | blue, 0 | blue, 0 | black, 1 | black, 1 |
|---|---|---|---|---|
| second drawn, $y$ | blue, 0 | black, 1 | blue, 0 | black, 1 |
| $f(x,y)$ | $\frac{8 \cdot 7}{14 \cdot 13} = \frac{28}{91}$ | $\frac{8 \cdot 6}{14 \cdot 13} = \frac{24}{91}$ | $\frac{6 \cdot 8}{14 \cdot 13} = \frac{24}{91}$ | $\frac{6 \cdot 5}{14 \cdot 13} = \frac{15}{91}$ |

(i) **True** (ii) **False**

(d) Chance of choosing a blue marble in *first* of two draws is
$f_X(0) = P(X=0) = f(0,0) + f(0,1) =$
(i) $\frac{\mathbf{28}}{\mathbf{91}} + \frac{\mathbf{28}}{\mathbf{91}}$ (ii) $\frac{\mathbf{28}}{\mathbf{91}} + \frac{\mathbf{24}}{\mathbf{91}}$ (iii) $\frac{\mathbf{28}}{\mathbf{91}} + \frac{\mathbf{15}}{\mathbf{91}}$ (iv) $\frac{\mathbf{24}}{\mathbf{91}} + \frac{\mathbf{15}}{\mathbf{91}}$.

(e) Chance of choosing a black marble in *first* of two draws is
$f_X(1) = P(X=1) = f(1,0) + f(1,1) =$
(i) $\frac{\mathbf{28}}{\mathbf{91}} + \frac{\mathbf{28}}{\mathbf{91}}$ (ii) $\frac{\mathbf{28}}{\mathbf{91}} + \frac{\mathbf{24}}{\mathbf{91}}$ (iii) $\frac{\mathbf{28}}{\mathbf{91}} + \frac{\mathbf{15}}{\mathbf{91}}$ (iv) $\frac{\mathbf{24}}{\mathbf{91}} + \frac{\mathbf{15}}{\mathbf{91}}$.

(f) $P(X+Y=1) = f(0,1) + f(1,0) = \frac{24}{91} + \frac{24}{91} =$
(i) $\frac{\mathbf{48}}{\mathbf{91}}$ (ii) $\frac{\mathbf{28}}{\mathbf{91}}$ (iii) $\frac{\mathbf{24}}{\mathbf{91}}$ (iv) $\frac{\mathbf{15}}{\mathbf{91}}$.

(g) The joint density, including the marginal probabilities,

| | $f(x,y)$ | $x$: blue, 0 | black, 1 | $f_Y(y) = P(Y=y)$ |
|---|---|---|---|---|
| $y$ | blue, 0 | $\frac{28}{91}$ | $\frac{24}{91}$ | $\frac{52}{91}$ |
| | black, 1 | $\frac{24}{91}$ | $\frac{15}{91}$ | $\frac{39}{91}$ |
| | $f_X(x) = P(X=x)$ | $\frac{52}{91}$ | $\frac{39}{91}$ | 1 |

(i) **True** (ii) **False**

(h) Are first draw $X$ and second draw $Y$ independent? Since

$$\begin{aligned} f(0,0) &= \frac{28}{91} \approx 0.307 &\neq& \quad f_X(0) \cdot f_Y(0) = \frac{52}{91} \cdot \frac{52}{91} \approx 0.327 \\ f(0,1) &= \frac{24}{91} \approx 0.264 &\neq& \quad f_X(0) \cdot f_Y(1) = \frac{52}{91} \cdot \frac{39}{91} \approx 0.245 \\ f(1,0) &= \frac{24}{91} \approx 0.264 &\neq& \quad f_X(1) \cdot f_Y(0) = \frac{39}{91} \cdot \frac{52}{91} \approx 0.245 \\ f(1,1) &= \frac{15}{91} \approx 0.165 &\neq& \quad f_X(1) \cdot f_Y(1) = \frac{39}{91} \cdot \frac{39}{91} \approx 0.184 \end{aligned}$$

$X$ and $Y$ are (i) **independent** (ii) **dependent**.

Only necessary to show *one* of four equations unequal to one another to demonstrate dependence; on the other hand, must show *all* equations to be equal to one another to show *in*dependence.

2. *Another discrete joint pmf.* Consider bivariate function

$$f(x,y) = \frac{xy}{18}, \quad x = 1,2; y = 1,2,3.$$

(a) This function is a pmf because $0 \leq f(x,y) \leq 1$, $x = 1,2; y = 1,2,3$ and

$$\sum_{x=1}^{2} \sum_{y=1}^{3} \frac{xy}{18} = \frac{1 \cdot 1}{18} + \frac{1 \cdot 2}{18} + \frac{1 \cdot 3}{18} + \frac{2 \cdot 1}{18} + \frac{2 \cdot 2}{18} + \frac{2 \cdot 3}{18} =$$

(i) **0** (ii) **0.5** (iii) **0.75** (iv) **1**.

(b) Marginal pmf of $X$ is

$$f_X(x) = \sum_{y=1}^{3} \frac{xy}{18} = \frac{x + 2x + 3x}{18} =$$

(i) $\boldsymbol{\frac{x}{3}}$ (ii) $\boldsymbol{\frac{x}{6}}$ (iii) $\boldsymbol{\frac{x}{9}}$ (iv) $\boldsymbol{\frac{x}{10}}$.

(c) Marginal pmf of $Y$ is

$$f_Y(y) = \sum_{x=1}^{2} \frac{xy}{18} = \frac{y + 2y}{18} =$$

(i) $\boldsymbol{\frac{y}{3}}$ (ii) $\boldsymbol{\frac{y}{6}}$ (iii) $\boldsymbol{\frac{y}{9}}$ (iv) $\boldsymbol{\frac{y}{10}}$.

(d) Since

$$f(x,y) = \frac{xy}{18} = f_X(x) \cdot f_Y(y) = \frac{x}{3} \cdot \frac{y}{6}$$

$X$ and $Y$ are (i) **independent** (ii) **dependent**.

(e) $P(X + Y \leq 3) = f(1,1) + f(1,2) + f(2,1) = \frac{1 \cdot 1}{18} + \frac{1 \cdot 2}{18} + \frac{2 \cdot 1}{18} =$
(i) $\boldsymbol{\frac{3}{18}}$ (ii) $\boldsymbol{\frac{4}{18}}$ (iii) $\boldsymbol{\frac{5}{18}}$ (iv) $\boldsymbol{\frac{6}{18}}$.

(f) $P\left(\frac{X}{Y} = 1\right) = f(1,1) + f(2,2) = \frac{1 \cdot 1}{18} + \frac{2 \cdot 2}{18} =$
(i) $\mathbf{\frac{3}{18}}$ (ii) $\mathbf{\frac{4}{18}}$ (iii) $\mathbf{\frac{5}{18}}$ (iv) $\mathbf{\frac{6}{18}}$.

3. *Continuous pdf: weight and amount of salt in potato chips.* Two machines fill potato chip bags. Although each bag should weigh 50 grams each and contain 5 milligrams of salt, in fact, because of differing machines, weight, $X$, and amount of salt, $Y$, placed in each bag varies according to two graphs below. Consider

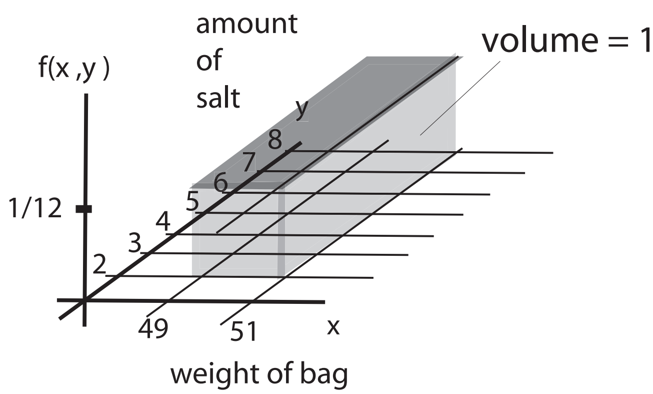



Figure 3.15: Continuous joint bivariate function: potato chips

following function for potato chip machine

$$f(x,y) = \begin{cases} \frac{1}{12}, & 49 \leq x \leq 51, 2 \leq y \leq 8 \\ 0 & \text{elsewhere} \end{cases}$$

(a) This function is a pdf because $49 \leq x \leq 51, 2 \leq y \leq 8$ and

$$\int_2^8 \int_{49}^{51} \frac{1}{12}\, dx\, dy = \frac{1}{12} \int_2^8 (x)_{x=49}^{x=51}\, dy = \frac{1}{12} \int_2^8 2\, dy = \frac{2}{12} (y)_{y=2}^{y=8} = \frac{2}{12} \cdot 6 =$$

(i) **0** (ii) **0.5** (iii) **0.75** (iv) **1**.

(b) Marginal for $X$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_2^8 \frac{1}{12}\, dy = \frac{1}{12}(y)_{y=2}^{y=8} =$$

(i) $\mathbf{\frac{1}{2}}$ (ii) $\mathbf{\frac{1}{3}}$ (iii) $\mathbf{\frac{1}{4}}$ (iv) $\mathbf{\frac{1}{5}}$,
where $49 \leq x \leq 51$.

(c) Marginal for $Y$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_{49}^{51} \frac{1}{12}\, dx = \frac{1}{12}(x)_{x=49}^{x=51} =$$

(i) $\mathbf{\frac{1}{3}}$ (ii) $\mathbf{\frac{1}{4}}$ (iii) $\mathbf{\frac{1}{5}}$ (iv) $\mathbf{\frac{1}{6}}$,
where $2 \leq y \leq 8$.

(d) *Independence?* Since

$$f(x,y) = \frac{1}{12} = f_1(x)f_2(y) = \frac{1}{2} \times \frac{1}{6},$$

random variables $X$ and $Y$ are (i) **dependent** (ii) **independent**

4. *Continuous Bivariate Function: What Is Constant k?*

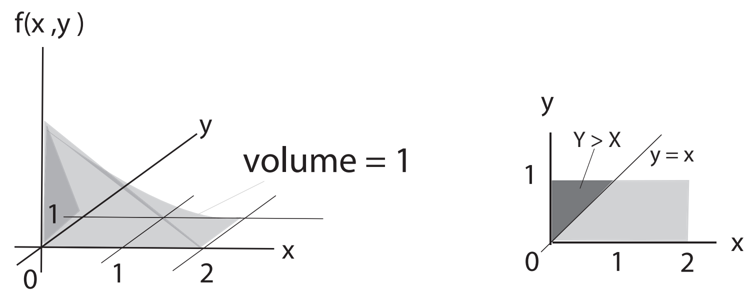



Figure 3.16: Continuous bivariate density $f(x,y) = (2-x)(1-y)$

(a) Determine $k$ so that

$$f(x,y) = \begin{cases} k(2-x)(1-y) & 0 \le x \le 2, 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

is a joint probability density function. Since

$$\begin{aligned}
\int_0^1 \int_0^2 k(2-x)(1-y)\,dx\,dy &= k\int_0^1 \int_0^2 (2-x-2y+xy)\,dx\,dy \\
&= k\int_0^1 \left(2x - \frac{1}{2}x^2 - 2xy + \frac{1}{2}x^2y\right)_{x=0}^{x=2} dy \\
&= k\int_0^1 (2-2y)\ dy \\
&= k\big(2y-y^2\big)_{y=0}^{y=1} = k\,(2-1) = 1,
\end{aligned}$$

so $k =$ (i) $\mathbf{\frac{1}{4}}$ (ii) $\mathbf{\frac{2}{4}}$ (iii) $\mathbf{\frac{3}{4}}$ (iv) **1**.

(b) Marginal $f_X(x)$.

$$\begin{aligned}
\int_0^1 (2-x)(1-y)\,dy &= \int_0^1 (2-x-2y+xy)\,dy \\
&= \left(2y - xy - y^2 + \frac{1}{2}xy^2\right)_{y=0}^{y=1} =
\end{aligned}$$

(i) $\mathbf{\frac{1}{2} - \frac{1}{2}\boldsymbol{x}}$ (ii) $\mathbf{1 - \frac{1}{2}\boldsymbol{x}}$ (iii) $\mathbf{\frac{3}{2} - \frac{1}{2}\boldsymbol{x}}$,
where $0 \leq x \leq 2$.

(c) Marginal $f_Y(y)$.

$$\begin{aligned}\int_0^2 (2-x)(1-y)\,dx &= \int_0^2 (2-x-2y+xy)\,dx \\ &= \left(2x - \frac{1}{2}x^2 - 2xy + \frac{1}{2}x^2 y\right)_{x=0}^{x=2} =\end{aligned}$$

(i) $\mathbf{2 - 2\boldsymbol{y}}$ (ii) $\mathbf{1 - \frac{1}{2}\boldsymbol{y}}$ (iii) $\mathbf{\frac{3}{2} - \frac{1}{2}\boldsymbol{y}}$,
where $0 \leq y \leq 1$.

(d) Since

$$f(x,y) = (2-x)(1-y) = f_X(x) f_Y(y) = \left(1 - \frac{1}{2}x\right) \times (2-2y),$$

random variables $X$ and $Y$ are (i) **dependent** (ii) **independent**

(e) Determine $P(Y > X)$

$$\begin{aligned}\int_0^1 \int_0^y (2-x)(1-y)\,dx\,dy &= \int_0^1 \int_0^y (2-x-2y+xy)\,dx\,dy \\ &= \int_0^1 \left(2x - \frac{1}{2}x^2 - 2xy + \frac{1}{2}x^2 y\right)_{x=0}^{x=y} dy \\ &= \int_0^1 \left(2y - \frac{5}{2}y^2 + \frac{1}{2}y^3\right) dy \\ &= \left(y^2 - \frac{5}{6}y^3 + \frac{1}{8}y^4\right)_{y=0}^{y=1} =\end{aligned}$$

(i) $\mathbf{\frac{6}{24}}$ (ii) $\mathbf{\frac{7}{24}}$ (iii) $\mathbf{\frac{8}{24}}$ (iv) **1**.

"horizontal" slices of dark region: as $y$ ranges from 0 to 1, $x$ ranges from 0 to $y$, the line $y = x$

## 3.7 Functions of Independent Random Variables

The *mean* or *expected value* of a *function* of two (continuous) random variables, $X$, $Y$, $u(X,Y)$, is

$$E[u(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x,y) f(x,y)\,dx\,dy$$

where

- $E[X+Y] = E[X] + E[Y]$, whether independent or not

- $E(XY) = E(X)E(Y)$ if and only if $X$ and $Y$ are independent
- Define *covariance* $\sigma_{XY}^2 = E(XY) - E(X)E(Y)$, then for $W = X + Y$,

$$Var(X+Y) = Var(W) = \sigma_W^2 = \begin{cases} \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} & \text{dependent} \\ \sigma_X^2 + \sigma_Y^2 & \text{independent} \end{cases}$$

- If $X$, $Y$ independent random variables, then for $W = aX + bY$, mgf

$$M_W(t) = M_X(at)M_Y(bt),$$

which implies if $X$ is $N(\mu_X, \sigma_X^2)$ and $Y$ is $N(\mu_Y, \sigma_Y^2)$, then

$$W = aX + bY \text{ is } N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2).$$

- If $X$ and $Y$ are independent, $Z = g(X)$ and $W = h(Y)$ are also independent, in particular, $Z = X^2$ and $W = Y^2$ are independent.

**Exercise 3.7 (Functions of Independent Random Variables)**

1. *Continuous Expected Value Calculations: Potato Chips.* Although each bag should weigh 50 grams each and contain 5 milligrams of salt, in fact, because of differing machines, weight and amount of salt placed in each bag varies according to the following joint pdf.

$$f(x,y) = \begin{cases} \frac{1}{12}, & 49 \leq x \leq 51, 2 \leq y \leq 8 \\ 0 & \text{elsewhere} \end{cases}$$

(a) The expected value of $u(x,y) = xy$, $E[XY]$, is

$$\begin{aligned} \int_2^8 \int_{49}^{51} (xy)\, f(x,y)\, dxdy &= \int_2^8 \int_{49}^{51} (xy)\, \frac{1}{12}\, dxdy = \frac{1}{12} \int_2^8 y \left( \int_{49}^{51} x\, dx \right) dy \\ &= \frac{1}{12} \int_2^8 y \left( \frac{1}{2} x^2 \right)_{x=49}^{x=51} dy = \frac{1}{12} \cdot \frac{1}{2}(51^2 - 49^2) \int_2^8 y\, dy \\ &= \frac{200}{24} \int_2^8 y\, dy = \frac{200}{24} \left( \frac{1}{2} y^2 \right)_{y=2}^{y=8} = \end{aligned}$$

(i) **5** (ii) **50** (iii) **55** (iii) **250**.

(b) The expected value of $u(x,y) = x$, $E[X]$, is

$$\begin{aligned} \int_2^8 \int_{49}^{51} x f(x,y)\, dxdy &= \int_2^8 \int_{49}^{51} x \frac{1}{12}\, dxdy = \frac{1}{12} \int_2^8 \left( \int_{49}^{51} x\, dx \right) dy \\ &= \frac{1}{12} \int_2^8 \left( \frac{1}{2} x^2 \right)_{x=49}^{x=51} dy = \frac{1}{12} \cdot \frac{1}{2}(51^2 - 49^2) \int_2^8 1\, dy \\ &= \frac{200}{24} \int_2^8 1\, dy = \frac{200}{24} (y)_{y=2}^{y=8} = \end{aligned}$$

(i) **5** (ii) **50** (iii) **55** (iii) **250**.

(c) The expected value of $u(x, y) = y$, $E[Y]$, is

$$\begin{aligned} \int_2^8 \int_{49}^{51} y f(x,y)\, dxdy &= \int_2^8 \int_{49}^{51} y \frac{1}{12}\, dxdy = \frac{1}{12} \int_2^8 y \left( \int_{49}^{51} 1\, dx \right) dy \\ &= \frac{1}{12} \int_2^8 y\, (x)_{x=49}^{x=51}\, dy = \frac{1}{12}(51-49) \int_2^8 y\, dy \\ &= \frac{2}{12} \int_2^8 y\, dy = \frac{1}{6} \left( \frac{1}{2} y^2 \right)_{y=2}^{y=8} = \end{aligned}$$

(i) **5** (ii) **50** (iii) **55** (iii) **250**.

(d) Since

$$E(XY) = 250 = 5 \cdot 50 = E(X) \cdot E(Y)$$

random variables $X$ and $Y$ are (i) **independent** (ii) **dependent**.

(e) Whether or not $X$ and $Y$ are independent,

$$E(X+Y) = E(X) + E(Y) = 50 + 5 =$$

(i) **5** (ii) **50** (iii) **55** (iii) **250**.

(f) Find covariance $\sigma_{XY}^2$.

$$\sigma_{XY}^2 = E(XY) - E(X)E(Y) = 250 - 5 \cdot 50 =$$

(i) **0** (ii) **28** (iii) **315** (iii) $\mathbf{\frac{90012}{36}}$.

(g) Find $E(X + 3Y + XY)$.

$$E(X + 3Y + XY) = E(X) + 3E(Y) + E(XY) = 50 + 3(5) + 250 =$$

(i) **0** (ii) **28** (iii) **315** (iii) $\mathbf{\frac{90012}{36}}$.

(h) Find $E(X^2)$.

$$\begin{aligned} \int_2^8 \int_{49}^{51} x^2 f(x,y)\, dxdy &= \int_2^8 \int_{49}^{51} x^2 \frac{1}{12}\, dxdy = \frac{1}{12} \int_2^8 \left( \int_{49}^{51} x^2\, dx \right) dy \\ &= \frac{1}{12} \int_2^8 \left( \frac{1}{3} x^3 \right)_{x=49}^{x=51} dy = \frac{1}{12} \cdot \frac{1}{3}(51^3 - 49^3) \int_2^8 1\, dy \\ &= \frac{15002}{36} \int_2^8 1\, dy = \frac{15002}{36}\, (y)_{y=2}^{y=8} = \end{aligned}$$

(i) **5** (ii) **28** (iii) **315** (iii) $\mathbf{\frac{90012}{36}}$.

(i) Find $Var(X) = \sigma_X^2$.

$$Var(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu_X^2 = \frac{90012}{36} - 50^2 =$$

(i) $\mathbf{\frac{1}{3}}$ (ii) **28** (iii) **315** (iii) $\mathbf{\frac{90012}{36}}$.

(j) Find $E\left[Y^2\right]$.

$$\begin{aligned}\int_2^8 \int_{49}^{51} y^2 f(x,y)\,dxdy &= \int_2^8 \int_{49}^{51} y^2 \frac{1}{12}\,dxdy = \frac{1}{12}\int_2^8 y^2 \left(\int_{49}^{51} 1\,dx\right) dy \\ &= \frac{1}{12}\int_2^8 y^2\,(x)_{x=49}^{x=51}\,dy = \frac{1}{12}(51-49)\int_2^8 y^2\,dy \\ &= \frac{2}{12}\int_2^8 y^2\,dy = \frac{1}{6}\left(\frac{1}{3}y^3\right)_{y=2}^{y=8} =\end{aligned}$$

(i) **5** (ii) **28** (iii) **315** (iii) $\mathbf{\frac{90012}{36}}$.

(k) Find $Var(Y) = \sigma_Y^2$.

$$Var(Y) = E(Y^2) - [E(Y)]^2 = E(Y^2) - \mu_Y^2 = 28 - 5^2 =$$

(i) $\mathbf{\frac{1}{3}}$ (ii) **3** (iii) **315** (iii) $\mathbf{\frac{90012}{36}}$.

(l) Find $Var(X+Y)$. Since $X$ and $Y$ are independent,

$$Var(X+Y) = Var(X) + Var(Y) = \sigma_X^2 + \sigma_Y^2 = \frac{1}{3} + 3 =$$

(i) $\mathbf{\frac{10}{3}}$ (ii) $\mathbf{\frac{11}{3}}$ (iii) $\mathbf{\frac{12}{3}}$ (iii) $\mathbf{\frac{13}{3}}$.

2. *Expected value for the sum of normal random variables.*
Assume $X$ is $N(\mu_X, \sigma_X^2) = N(2, 2^2)$ and $Y$ is $N(-1, 4^2)$ and let $W = 4X + 5Y$.

   (a) What is distribution of $W$?

$$N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2) = N(4(2) + 5(-1), 4^2(2^2) + 5^2(4^2)) =$$

   (i) $\boldsymbol{N(2, 412)}$ (ii) $\boldsymbol{N(3, 464)}$ (iii) $\boldsymbol{N(4, 426)}$ (iii) $\boldsymbol{N(5, 428)}$.

   (b) So $P(W < 4) \approx$ (i) **0.4935** (ii) **0.5185** (iii) **0.5199** (iii) **0.6001**

```
pnorm(4,3,sqrt(464)) # normal P(X < 4), m = 3, sd = sqrt(416)
[1] 0.5185138
```

   Or, using table C.1,

$$P(W < 4) = P\left(Z < \frac{4-3}{\sqrt{416}}\right) \approx P(Z < 0.05) \approx$$

   (i) **0.4935** (ii) **0.5196** (iii) **0.5199** (iii) **0.6001**

3. *Expected number of matches.* Ten people throw ten tickets with their names on each ticket into a jar, then draw one ticket out of the jar at random (and put it back in the jar). Let $X$ be the number of people who select their own ticket out of the jar. Let

$$X = X_1 + X_2 + \cdots + X_{10}$$

where

$$X_i = \begin{cases} 1 & \text{if } i\text{th person selects own ticket} \\ 0 & \text{if } i\text{th person does not select their own ticket} \end{cases}$$

(a) Since each person chooses any of the ten tickets with equal chance, $E[X_i] = 1 \times \frac{1}{10} + 0 \times \frac{9}{10} =$ (i) $\mathbf{\frac{1}{10}}$ (ii) $\mathbf{\frac{2}{10}}$ (iii) $\mathbf{\frac{3}{10}}$.

(b) So expected number of ten individuals to choose their own ticket is $E(X) = E(X_1) + \cdots + E(X_{10}) = 10 \times \frac{1}{10} =$ (i) $\mathbf{\frac{8}{10}}$ (ii) $\mathbf{\frac{9}{10}}$ (iii) $\mathbf{\frac{10}{10}}$. We would expect one of ten individuals to choose their own ticket.

(c) *If* $n$ individuals played this game, then we would expect $E(X) = E(X) + \cdots + E(Y_n) = n\left(\frac{1}{n}\right) =$ (i) $\boldsymbol{\frac{n-1}{n}}$ (ii) $\boldsymbol{\frac{n}{n}}$ (iii) $\boldsymbol{\frac{n+1}{n}}$. Again, we would expect one of $n$ individuals to choose their own ticket.

## 3.8 Central Limit Theorem

A *population* is a set of measurements or observations of a collection of objects. A *sample* is a selected subset of a population. A *parameter* is a numerical quantity calculated from a population, whereas a *statistic* is a numerical quantity calculated from a sample. The population is assumed modelled by some random variable $X$ with probability distribution, for example, the normal distribution, $N(\mu, \sigma^2)$, with population parameter mean $\mu$ and population parameter variance $\sigma^2$. A typical example of a sample statistic is the sample mean of $n$ of the $X$ random variables,

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

If $X_1, X_2, \ldots, X_n$ are mutually independent random variables where each is $N(\mu, \sigma^2)$, then $\bar{X}_n$ is

$$N(\mu_{\bar{X}_n}, \sigma^2_{\bar{X}_n}) = N(\mu, \frac{\sigma^2}{n}).$$

In fact, the *central limit theorem* (CLT) says if $X_1, X_2, \ldots, X_n$ are mutually independent random variables where each which common $\mu$ and $\sigma^2$, then as $n \to \infty$,

$$\bar{X}_n \to N\left(\mu, \frac{\sigma^2}{n}\right),$$

*no matter what the distribution of the population.* Often $n \geq 30$ is "large enough" for the CLT to apply.

**Exercise 3.8 (Central Limit Theorem)**

1. *Practice with CLT: average, $\bar{X}$.*

   (a) *Number of burgers.*
   Number of burgers, $X$, made per minute at Best Burger averages $\mu_X = 2.7$ burgers with a standard deviation of $\sigma_X = 0.64$ of a burger. Consider average number of burgers made over random $n = 35$ minutes during day.

i. $\mu_{\bar{X}} = \mu_X =$ (i) **2.7** (ii) **2.8** (iii) **2.9**.

ii. $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.64}{\sqrt{35}} =$ (i) **0.10817975** (ii) **0.1110032** (iii) **0.13099923**.

iii. $P\left(\bar{X} > 2.75\right) \approx$ (i) **0.30** (ii) **0.32** (iii) **0.35**.

```
1 - pnorm(2.75,2.7,0.64/sqrt(35)) # normal P(X > 2.75) = 1 - P(X < 2.75)
[1] 0.3219712
```

iv. $P\left(2.65 < \bar{X} < 2.75\right) = P\left(\bar{X} < 2.75\right) - P\left(\bar{X} < 2.65\right) \approx$
(i) **0.36** (ii) **0.39** (iii) **0.45**.

```
pnorm(2.75,2.7,0.64/sqrt(35)) -  pnorm(2.65,2.7,0.64/sqrt(35))
[1] 0.3560576
```

(b) *Temperatures.*
Temperature, $X$, on any given day during winter in Laporte averages $\mu_X = 0$ degrees with standard deviation of $\sigma_X = 1$ degree. Consider average temperature over random $n = 40$ days during winter.

i. $\mu_{\bar{X}} = \mu_X =$ (i) **0** (ii) **1** (iii) **2**.

ii. $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1}{\sqrt{40}} =$ (i) **0.0900234** (ii) **0.15811388** (iii) **0.23198455**.

iii. $P\left(\bar{X} > 0.2\right) \approx$ (i) **0.03** (ii) **0.10** (iii) **0.15**.

```
1 - pnorm(0.2,0,1/sqrt(40)) # normal P(X > 0.2) = 1 - P(X < 0.2)
[1] 0.1029516
```

iv. $P\left(\bar{X} > 0.3\right) \approx$ (i) **0.03** (ii) **0.10** (iii) **0.15**.

```
1 - pnorm(0.3,0,1/sqrt(40)) # normal P(X > 0.3) = 1 - P(X < 0.3)
[1] 0.02888979
```

Since $P\left(\bar{X} > 0.3\right) \approx 0.03 < 0.05$, $0.3^o$ (i) **is** (ii) **is not** unusual.

(c) *Another example.*
Suppose $X$ has distribution where $\mu_X = 1.7$ and $\sigma_X = 1.5$.

i. $\mu_{\bar{X}} = \mu_X =$ (i) **2.3** (ii) **1.7** (iii) **2.4**.

ii. $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1.5}{\sqrt{49}} =$ (i) **0.0243892** (ii) **0.14444398** (iii) **0.21428572**.

iii. If $n = 49$, $P(-2 < \bar{X} < 2.75) \approx$ (i) **0.58** (ii) **0.86** (iii) **0.999**.

```
pnorm(2.75,1.7,1.5/sqrt(49)) -  pnorm(-2,1.7,1.5/sqrt(49)) #  P(X-bar < 2.75) - P(X-bar < -2)
[1] 0.9999995
```

iv. **True** (ii) **False**.
If $n = 15$, $P(-2 < \bar{X} < 2.75)$ cannot be calculated since $n = 15 < 30$.

v. $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1.5}{\sqrt{15}} =$ (i) **0.0243892** (ii) **0.14444398** (iii) **0.38729835**.

vi. If $n = 15$ *and* normal,
$P(-2 < \bar{X} < 2.75) \approx$ (i) **0.75** (ii) **0.78** (iii) **0.997**.

```
pnorm(2.75,1.7,1.5/sqrt(15)) -  pnorm(-2,1.7,1.5/sqrt(15)) #  P(X-bar < 2.75) - P(X-bar < -2)
[1] 0.9966469
```

(d) *Dice average.*
What is the chance, in $n = 30$ rolls of a fair die, average is between 3.3 and 3.7, $P\left(3.3 < \bar{X} < 3.7\right)$? What if $n = 3$?

i. $\mu_{\bar{X}} = \mu_X = 1\left(\frac{1}{6}\right) + \cdots + 6\left(\frac{1}{6}\right) =$ (i) **2.3** (ii) **3.5** (iii) **4.3**.

```
x <- 1:6 # values of die
px <- c(1/6,1/6,1/6,1/6,1/6,1/6) # probabilities: 1/6 repeated 6 times
EX <- sum(x*px); EX # E(X)
[1] 3.5
```

ii. $\sigma_X = \sqrt{(1-3.5)^2\left(\frac{1}{6}\right) + \cdots + (6-3.5)^2\left(\frac{1}{6}\right)} =$
(i) **1.7078252** (ii) **2.131145** (iii) **3.3409334**.

```
SDX <- sqrt(sum((x-EX)^2*px)); SDX
[1] 1.707825
```

iii. If $n = 30$, $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1.71}{\sqrt{30}} =$ (i) **0.31** (ii) **0.75** (iii) **1.14**.

iv. If $n = 30$, $P\left(3.3 < \bar{X} < 3.7\right) \approx$ (i) **0** (ii) **0.20** (iii) **0.48**.

```
pnorm(3.7,EX,SDX/sqrt(30)) -  pnorm(3.3,EX,SDX/sqrt(30)) #  P(X-bar < 3.7) - P(X-bar < 3.3)
[1] 0.4787547
```

v. (i) **True** (ii) **False**.
If $n = 3$, $P\left(3.3 < \bar{X} < 3.7\right)$ cannot be calculated because $n = 3 < 30$.

2. *Understanding CLT: Montana fishing trip.*

(a) *Sampling distributions of average, $n = 1, 2, 3$.* As random sample size, $n$, increases, sampling distribution of average, $\bar{X}$, changes shape and becomes more

i. rectangular–shaped.
ii. bell–shaped.
iii. triangular–shaped.

*Central limit theorem (CLT)* says *no matter what the original parent distribution*, sampling distribution of average is typically normal when $n > 30$.

(b) In addition to sampling distribution becoming more normal–shaped as random sample size increases, mean of average, $\mu_{\bar{X}} = 1.8$

i. decreases and is equal to $\frac{\sigma_X^2}{n}$,
ii. remains same and is equal to $\mu_X = 1.8$,
iii. increases and is equal to $n\mu_X$,

and standard deviation of average, $\sigma_{\bar{X}}$

i. decreases and is equal to $\frac{\sigma_X}{\sqrt{n}}$.
ii. remains same and is equal to $\sigma_X$.
iii. increases and is equal to $n\sigma_X$.

(c) After $n = 30$ trips to lake, sampling distribution in average number of fish caught is essentially *normal* (why?) where
$\mu_{\bar{X}} = \mu_X =$ (i) **1.2** (ii) **1.5** (iii) **1.8**,
$\sigma_{\bar{X}} = \frac{0.75}{\sqrt{30}} \approx$ (i) **0.12677313** (ii) **0.13693064** (iii) **0.2449987**,
and chance average number of fish is *less than* 1.95 is
$P(\bar{X} < 1.95) \approx$ (i) **0.73** (ii) **0.86** (iii) **0.94**.

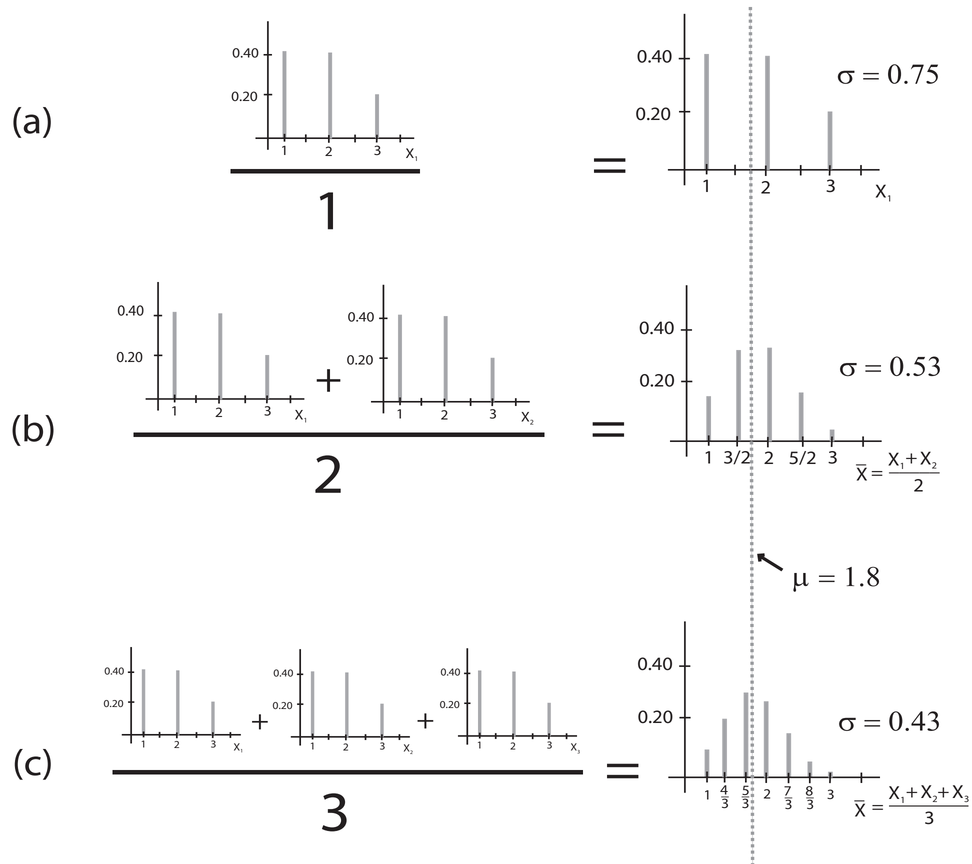



Figure 3.17: Comparing sampling distributions of sample mean

```
pnorm(1.95,1.8,0.75/sqrt(30)) # normal P(X-bar < 1.95)
```

```
[1] 0.8633392
```

(d) After $n = 35$ trips to lake, sampling distribution in average number of fish caught is essentially normal where
$\mu_{\bar{X}} = \mu_X =$ (i) **1.2** (ii) **1.5** (iii) **1.8**,
$\sigma_{\bar{X}} = \frac{0.75}{\sqrt{35}} \approx$ **0.12677313** (ii) **0.13693064** (ii) **0.2449987**,
and chance average number of fish is *less than* 1.95 is
$P(\bar{X} < 1.95) \approx$ (i) **0.73** (ii) **0.88** (iii) **0.94**.

```
pnorm(1.95,1.8,0.75/sqrt(35)) # normal P(X-bar < 1.95)
```

```
[1] 0.8816382
```

(e) Chance average number of fish is less than 1.95 after *30 trips*, $P(\bar{X} < 1.95) \approx 0.86$, is **smaller than** / **larger than** chance average number of fish is less than 1.95 after *35 trips*, $P(\bar{X} < 1.95) \approx 0.88$.

(f) The CLT is useful because (circle *one or more*):

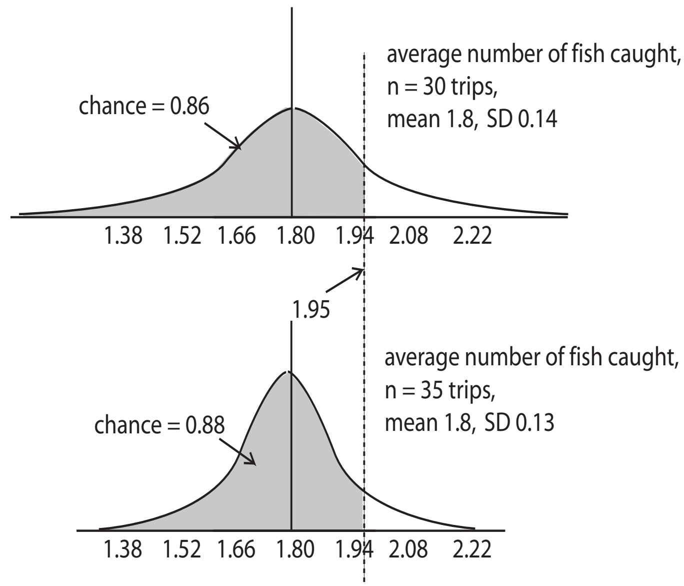



Figure 3.18: Chance when $n = 30$ compared to chance when $n = 35$

i. No matter what original parent distribution is, as long as a large enough random sample is taken, average of this sample follows a normal distribution.

ii. In practical situations where it is not known what parent probability distribution to use, as long as a large enough random sample is taken, average of this sample follows a normal distribution.

iii. Rather than having to deal with many different probability distributions, as long as a large enough random sample is taken, average of this sample follows *one* distribution, normal distribution.

iv. Many distributions in statistics rely in one way or another on normal distribution because of CLT.

(g) (i) **True** (ii) **False** Central limit theorem requires not only $n \geq 30$, but also a *random sample* of size $n \geq 30$ is used.

## 3.9 The Gamma and Related Distributions

Four related distributions which are important for statistics are discussed, including the gamma, chi-square, Student-$t$ and $F$ distributions.

- *Gamma distribution.*

- $X$ has *gamma distribution* with parameters $r$ and $\lambda, \lambda > 0$ with pdf

$$f(x) = \begin{cases} \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, & x > 0, \ r \text{ positive integer}, \\ \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0, \ r > 0 \end{cases}$$

  where *gamma function* $\Gamma(r)$, a *generalized factorial*, is

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} \, dy, \quad r > 0,$$

  where $r$ is *shape* parameter, $\lambda$ is *rate* parameter, and $\frac{1}{\lambda}$ is *scale* parameter

- $\mu = E(X) = \frac{r}{\lambda}, \quad \sigma^2 = Var(X) = \frac{r}{\lambda^2}, \quad M(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}, \quad t < \lambda$
- If $r$ is a positive integer and if *number* of events in an interval is Poisson with parameter $\lambda$, then the *time* until $r$th occurrence could be thought of as a gamma distribution with parameters $r$ and $\lambda$.

• *Chi-square distribution.*

- $X$ has *chi-square distribution with* $n > 0$ *degrees of freedom* and with pdf

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ x > 0,$$

  is a special case of the gamma distribution where $\lambda = \frac{1}{2}$ and $r = \frac{n}{2}$, and is described "$X$ is $\chi^2(n)$"

- $\mu = E(X) = n, \quad \sigma^2 = Var(X) = 2n, \quad M(t) = (1-2t)^{-\frac{n}{2}}, \ t < \frac{1}{2}$
- if $Z_1, \ldots Z_n$ each independent $N(0,1)$, $X = Z_1^2 + \cdots + Z_n^2$ is $\chi^2(n)$
- *critical value* $\chi_p^2(n)$ is a positive number where

$$P(X \geq \chi_p^2(n)) = p, \quad 0 \leq p \leq 1.$$

• *Student-t distribution.*

- $T$ has *Student-t distribution with* $n > 0$ *degrees of freedom* and with pdf

$$f(t) = \frac{\Gamma\left[\frac{n+1}{2}\right]}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}, \ -\infty < t < \infty.$$

- $\mu = E(T) = 0, \quad \sigma^2 = Var(T) = \frac{n}{n-2}, \quad M(t)$ is undefined.
- if $Z$ and $W$ are independent where $Z$ is $N(0,1)$ and $W$ is $\chi^2(n)$, then $T = \frac{Z}{\sqrt{\frac{W}{n}}}$ is Student-$t$ with $n$ degrees of freedom.

- ◦ *critical t-value* is a number $t_p(n)$ where

$$P(T \geq t_p(n)) = p, \quad 0 \leq p \leq 1.$$

- *F distribution.*
  - ◦ $X$ has *F distribution with n (numerator) and d (denominator) df* and with pdf

$$f(x) = \frac{\Gamma\left[\frac{n+1}{2}\right] n^{\frac{n}{2}} d^{\frac{d}{2}} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)(d+nx)^{\frac{n+d}{2}}}, \ x > 0.$$

  - ◦ $\mu = \frac{d}{d-2}$, $d > 2$, $\quad \sigma^2 = \frac{2d^2(n+d-2)}{n(d-2)^2(d-4)}$, $d > 4$, $\quad M(t)$ is undefined
  - ◦ if $U$ and $V$ are independent where $U$ is $\chi^2(n)$ and $V$ is $\chi^2(d)$, then $F = \frac{\frac{U}{n}}{\frac{V}{d}} = \frac{Ud}{Vn}$ has $F$ distribution with $n$ and $d$ degrees of freedom.

**Exercise 3.9 (The Gamma and Related Distributions)**

1. *Gamma distribution.*

   (a) *Gamma function,* $\Gamma(r)$

   i. $\Gamma(1.2) = \int_0^\infty y^{1.2-1} e^{-y}\, dy =$
   (i) **0.92** (ii) **1.12** (iii) **2.34** (iv) **2.67**.

```
gamma(1.2) # gamma function at 1.2
```

```
[1] 0.9181687
```

   ii. $\Gamma(2.2) \approx$
   (i) **0.89** (ii) **1.11** (iii) **1.84** (iv) **2.27**.

```
gamma(2.2) # gamma function
```

```
[1] 1.101802
```

   iii. $\Gamma(1) =$ (i) **0** (ii) **0.5** (iii) **0.7** (iv) **1**.

```
n <- c(1,2,3,4)
gamma(n) # gamma for vector of values: 1,2,3,4
```

```
[1] 1 1 2 6
```

   iv. $\Gamma(2) = (2-1)\Gamma(2-1) =$ (i) **0** (ii) **0.5** (iii) **0.7** (iv) **1**.

   v. $\Gamma(3) = 2\Gamma(2) =$ (i) **1** (ii) **2!** (iii) **3!** (iv) **4!**.

   vi. $\Gamma(4) = 3\Gamma(3) =$ (i) **1** (ii) **2!** (iii) **3!** (iv) **4!**.

   vii. In general, if $r = n$ is a positive integer,

$$\Gamma(n) = (n-1)!$$

   (i) **True** (ii) **False**

   (b) *Graphs of gamma density.*

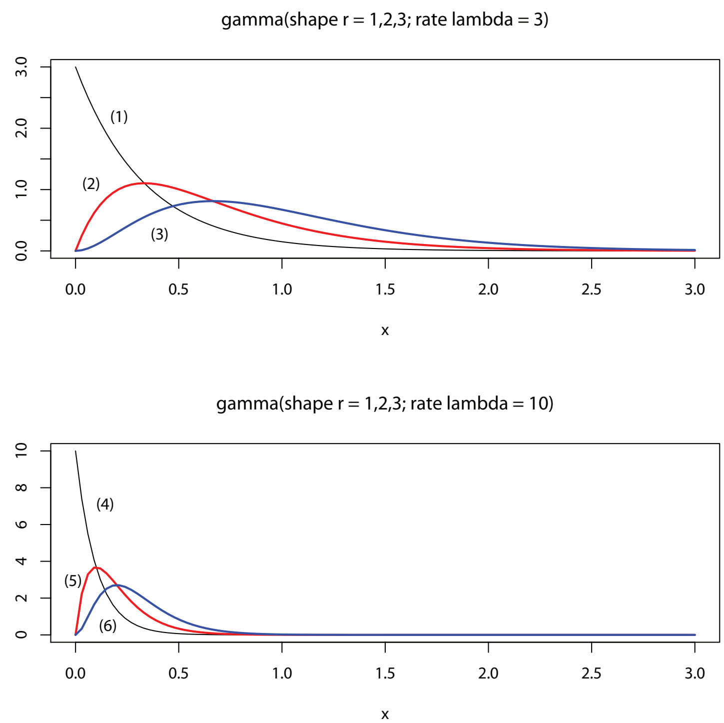



Figure 3.19: Gamma densities

i. Match gamma density, $(r, \lambda)$, to graph, (1) to (6).

| $(r, \lambda) =$ | $(1,3)$ | $(2,3)$ | $(3,3)$ | $(1,10)$ | $(2,10)$ | $(3,10)$ |
|---|---|---|---|---|---|---|
| graph | (1) | | | | | |

```
library(graphics)
par(mfrow = c(2,1))
plot(function(x) dgamma(x,1,3), 0, 3,
     main = "gamma(shape r = 1,2,3; rate lambda = 3)")
curve(dgamma(x,2,3), add = TRUE, col = "red", lwd = 2)
curve(dgamma(x,3,3), add = TRUE, col = "blue", lwd = 2)
plot(function(x) dgamma(x,1,10), 0, 3,
     main = "gamma(shape r = 1,2,3; rate lambda = 10)")
curve(dgamma(x,2,10), add = TRUE, col = "red", lwd = 2)
curve(dgamma(x,3,10), add = TRUE, col = "blue", lwd = 2)
par(mfrow = c(1,1))
```

ii. As $r$ increases, "center" (mean, $\mu = \frac{r}{\lambda}$) of gamma density
(i) **decreases.**
(ii) **remains the same.**
(iii) **increases.**

iii. As $\lambda$ increases, "dispersion" (variance, $\sigma^2 = \frac{r}{\lambda^2}$) of gamma density
(i) **decreases.**
(ii) **remains the same.**

(iii) **increases.**

(c) *Gamma distribution: area under gamma density.*

i. If $(r, \lambda) = (1, 3)$, $P(X < 1.3) = F(1.3) \approx$
(i) **0.59** (ii) **0.80** (iii) **0.81** (iv) **0.98**.

```
pgamma(1.3,1,3) # gamma, P(X < 1.3), r = 1, lambda = 3
[1] 0.9797581
```

ii. If $(r, \lambda) = (3, 3)$, $P(X > 0.5) = 1 - P(X \leq 0.5) = 1 - F(0.5) \approx$
(i) **0.59** (ii) **0.80** (iii) **0.81** (iv) **0.98**.

```
1 - pgamma(0.5,3,3) # gamma, P(X > 0.5), r = 3, lambda = 3
[1] 0.8088468
```

iii. If $(r, \lambda) = (3, 10)$, $P(0.5 < X < 0.1) = P(X \leq 0.5) - P(X \leq 0.1) \approx$
(i) **0.59** (ii) **0.80** (iii) **0.81** (iv) **0.98**.

```
pgamma(0.5,3,10) - pgamma(0.1,3,10) # gamma, P(0.1 < X < 0.5), r = 3, lambda = 10
[1] 0.7950466
```

(d) *Mean, variance and standard deviation of gamma distribution.*

i. If $(r, \lambda) = (2, 5)$, $\mu = \frac{r}{\lambda} = \frac{2}{5} =$
(i) **0.40** (ii) **0.08** (iii) **0.25** (iv) **0.28**.

ii. If $(r, \lambda) = (1.2, 4.3)$, $\mu = \frac{r}{\lambda} = \frac{1.2}{4.3} \approx$
(i) **0.40** (ii) **0.08** (iii) **0.25** (iv) **0.28**.

iii. If $(r, \lambda) = (2, 5)$, $\sigma^2 = \frac{r}{\lambda^2} = \frac{2}{5^2} \approx$
(i) **0.40** (ii) **0.08** (iii) **0.25** (iv) **0.28**.

iv. If $(r, \lambda) = (1.2, 4.3)$, $\sigma = \sqrt{\frac{r}{\lambda^2}} = \sqrt{\frac{2}{5^2}} \approx$
(i) **0.40** (ii) **0.08** (iii) **0.25** (iv) **0.28**.

2. *Gamma distribution again: time to fix car.* Assume the time, $X$, to fix a car is approximately a gamma with mean $\mu = 2$ hours and variance $\sigma^2 = 2$ hours$^2$.

(a) *What are $r$ and $\lambda$?* Since

$$\mu = \frac{r}{\lambda} = 2, \quad \sigma^2 = \frac{r}{\lambda^2} = \frac{r}{\lambda} \cdot \frac{1}{\lambda} = \mu \cdot \frac{1}{\lambda} = 2 \cdot \frac{1}{\lambda} = 2,$$

then $\lambda = 1$ and also $r = \mu\lambda = 2 \cdot 1 =$
(i) **1** (ii) **2** (iii) **3** (iv) **4**.

(b) In this case, the gamma density,

$$f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

is given by
(i)

$$f(x) = \begin{cases} xe^{-x}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

(ii)
$$f(x) = \begin{cases} \frac{xe^{-x}}{\Gamma(3)}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

(iii)
$$f(x) = \begin{cases} \frac{x^2 e^{-x/2}}{2^2\Gamma(1)}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

(c) What is the chance of waiting *at most* 4.5 hours?
Since $(r, \lambda) = (1, 2)$, $P(X < 4.5) = F(4.5) \approx$
(i) **0.002** (ii) **1.000** (iii) **0.870** (iv) **1.151**.

```
pgamma(4.5,1,2) # gamma, P(X < 4.5), r = 1, lambda = 2
```

```
[1] 0.9998766
```

(d) $P(X > 3.1) = 1 - P(X \leq 3.1) = 1 - F(3.1) \approx$
(i) **0.002** (ii) **1.000** (iii) **0.870** (iv) **1.151**.

```
1 - pgamma(3.1,1,2) # gamma, P(X > 3.1), r = 1, lambda = 2
```

```
[1] 0.002029431
```

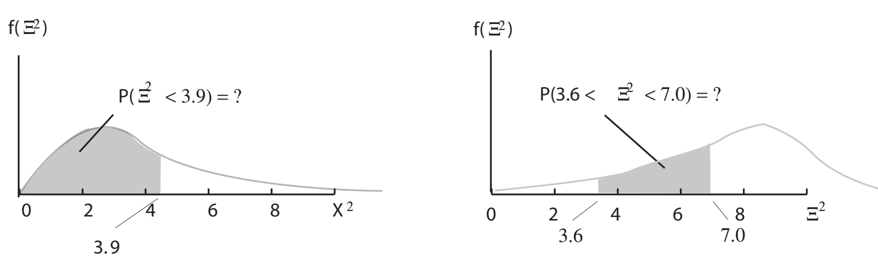
(e) What is the 90th percentile waiting time; in other words, what is that time such that 90% of waiting times are less than this time?
If $P(X < x) = 0.90$, then $x \approx$
(i) **0.002** (ii) **1.000** (iii) **0.870** (iv) **1.151**.

```
qgamma(0.90,1,2) # 90th percentile, r = 1, lambda = 2
```

```
[1] 1.151293
```

3. *Chi-square distribution: waiting time to order.* At McDonalds in Westville, waiting time to order (in minutes), $X$, follows a chi–square distribution.

(a) *Probabilities.*




(a) Chi-Square with 4 degrees of freedom (b) Chi-Square with 10 degrees of freedom

Figure 3.20: Chi–square probabilities

i. If $n = 4$, the probability of waiting less than 3.9 minutes is $P(X < 3.9) = F(3.9) \approx$
(i) **0.35** (ii) **0.45** (iii) **0.58** (iv) **0.66**.

```
pchisq(3.9,4) # chi-square, n = 4
[1] 0.5802915
```

ii. If $n = 10$, $P(3.6 < X < 7.0) \approx$
(i) **0.24** (ii) **0.33** (iii) **0.42** (iv) **0.56**.

```
pchisq(7,10) - pchisq(3.6,10) # chi-square, n = 10
[1] 0.2381484
```

iii. Chance of waiting time *exactly* 3 minutes, say, is *zero*, $P(X = 3) = 0$.
(i) **True** (ii) **False**

(b) *Critical value* $\chi^2_p(n)$ *(percentile).*

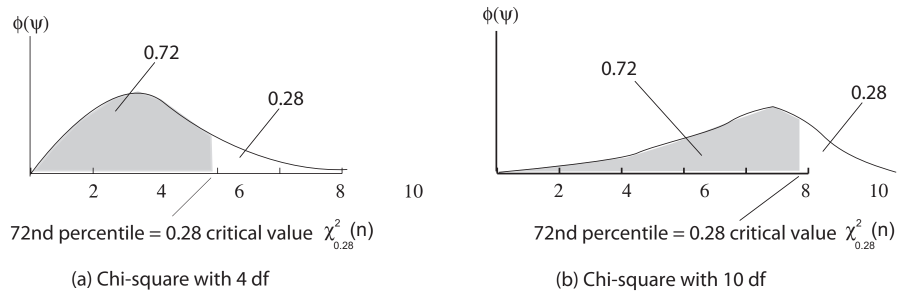



Figure 3.21: Chi–square percentiles

i. If $n = 4$ and $P(X > \chi^2_{0.28}(4)) = 0.28$, then 0.28 critical value $\chi^2_{0.28}(4) \approx$ (i) **3.1** (ii) **5.1** (iii) **8.3** (iv) **9.1**.

```
qchisq(0.28,4,lower.tail=FALSE) # chi-square, n = 4, 0.28 critical value
[1] 5.071894
```

ii. If $n = 4$ and $P(X < \chi^2_{0.28}(4)) = 0.72$, then 72nd percentile $\chi^2_{0.28}(4) \approx$ (i) **3.1** (ii) **5.1** (iii) **8.3** (iv) **9.1**.

```
qchisq(0.72,4,lower.tail=TRUE) # chi-square, n = 4, 72nd percentile
[1] 5.071894
```

iii. If $n = 10$ and $P(X > \chi^2_{0.28}(10)) = 0.28$, then $\chi^2_{0.28}(10) \approx$ (i) **2.5** (ii) **10.5** (iii) **12.1** (iv) **20.4**.

```
qchisq(0.28,10,lower.tail=FALSE) # chi-square, n = 10, 0.28 critical value
[1] 12.07604
```

iv. The 0.05 critical value for a chi-square with $n = 18$ df, is $\chi^2_{0.05}(18) \approx$ (i) **2.5** (ii) **10.5** (iii) **28.870** (iv) **28.869**.

```
qchisq(0.05,18,lower.tail=FALSE) # chi-square, n = 18, 0.05 critical value, 95th percentile
[1] 28.8693
```

or, equivalently using Table C.3
$\chi^2_{0.05}(18) \approx$ (i) **2.5** (ii) **10.5** (iii) **28.870** (iv) **28.869**.
Table C.4 can only be used for a restricted set of $(n, p)$.

v. The 0.05 critical value (95th percentile) is that waiting time such that 95% of the waiting times are less than this waiting time and 5% are more than this time. (i) **True** (ii) **False**

4. *Chi–square distribution again.*

   (a) If $n = 3$, the chi–square density,

   $$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

   is given by
   (i)

   $$f(x) = \begin{cases} \frac{1}{2} e^{-\frac{x}{2\Gamma(2)}}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

   (ii)

   $$f(x) = \begin{cases} \frac{1}{2\Gamma(2)} e^{-\frac{x}{2}}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

   (iii)

   $$f(x) = \begin{cases} \frac{1}{\Gamma(\frac{3}{2})2^{\frac{3}{2}}} x^{\frac{1}{2}} e^{-\frac{x}{2}}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

   (b) If $n = 3$, $\mu = E(Y) = n =$
   (i) **3** (ii) **4** (iii) **5** (iv) **6**.

   (c) If $n = 3$, $\sigma^2 = V(Y) = 2n =$
   (i) **3** (ii) **4** (iii) **5** (iv) **6**.

   (d) A chi–square with $n = 3$ degrees of freedom is a gamma with parameters $(r, \lambda) = \left(\frac{n}{2}, \frac{1}{2}\right) =$ (i) $\left(\mathbf{\frac{0}{2}, 2}\right)$ (ii) $\left(\mathbf{\frac{1}{2}, 2}\right)$ (iii) $\left(\mathbf{\frac{2}{2}, 2}\right)$ (iv) $\left(\mathbf{\frac{3}{2}, \frac{1}{2}}\right)$.

5. *Student-t distribution: temperatures in Westville.*

   Suppose temperature, $T$, on any given day during winter in Westville can be modelled as a Student-$t$ distribution with 4 degrees of freedom.

   (a) $P(T < 1.42) =$ (i) **0.886** (ii) **0.892** (iii) **0.945** (iv) **0.971**

   ```
   pt(1.42,4) # Student-t, 4 df, P(T < 1.42)
   [1] 0.8856849
   ```

   (b) $P(T < -2.11) =$ (i) **0.021** (ii) **0.031** (iii) **0.041** (iv) **0.051**

   ```
   pt(-2.11,4) # Student-t, 4 df, P(T < -2.11)
   [1] 0.05124523
   ```

   (c) $P(T > 0.54) =$ (i) **0.265** (ii) **0.295** (iii) **0.309** (iv) **0.351**

   ```
   1 - pt(0.54,4) # Student-t, 4 df, P(T > 0.54)
   [1] 0.3089285
   ```

   (d) $P(-1.73 < T < 1.62) =$ (i) **0.830** (ii) **0.876** (iii) **0.910** (iv) **0.992**

   ```
   pt(1.62,4) - pt(-1.73,4) # Student-t, 4 df, P(-1.73 < T < 1.62)
   ```

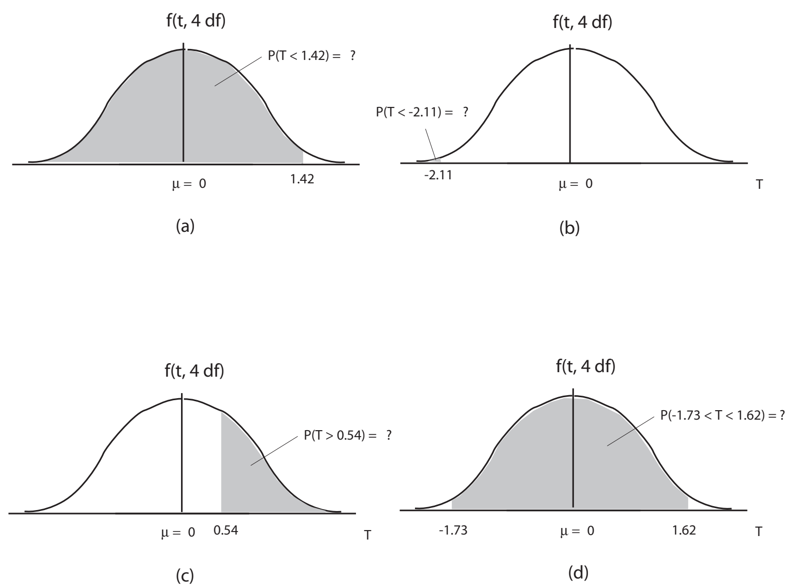



Figure 3.22: Student-$t$ distributions

```
[1] 0.8303853
```

(e) Compare $P(-1.73 < T < 1.62)$ for Student-$t$ when degrees of freedom $= 4, 24, 124$ and $N(0,1)$. Fill in the blanks.

| $t$, df $= 4$ | $t$, df $= 24$ | $t$, df $= 124$ | $N(0,1)$ |
|---|---|---|---|
| 0.8304 | | | |

The larger the degrees of freedom (implying larger sample size, $n$), the less flat the Student-$t$ distribution becomes, the more like the standard normal it becomes.

```
pt(1.62,4) - pt(-1.73,4) # Student-t, 4 df, P(-1.73 < T < 1.62)
pt(1.62,24) - pt(-1.73,24) # Student-t, 24 df, P(-1.73 < T < 1.62)
pt(1.62,124) - pt(-1.73,124) # Student-t, 124 df, P(-1.73 < T < 1.62)
pnorm(1.62) - pnorm(-1.73) # N(0,1), P(-1.73 < T < 1.62)
```

```
[1] 0.8303853
[1] 0.8926144
[1] 0.9030545
[1] 0.9055687
```

6. *Critical value $t_p(n)$ (percentile): temperatures.*

Again suppose temperature, $T$, on any given day during winter in Westville can be modelled as a Student-$t$ distribution with 4 degrees of freedom.

(a) The 92nd percentile (0.08 critical value $t_{0.08}(4)$) is
$t_{0.08}(4) =$ (i) $\mathbf{1.03^o}$ (ii) $\mathbf{1.32^o}$ (iii) $\mathbf{1.52^o}$ (iv) $\mathbf{1.72^o}$

```
qt(0.92,4) # Student-t, 4 df, 92nd percentile
```

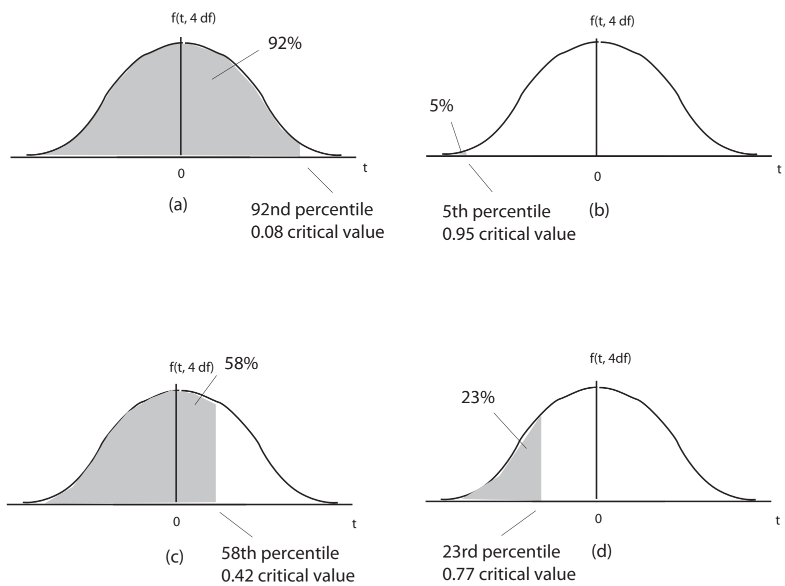



Figure 3.23: Critical values $t_p(n)$ for Student-$t$ distributions

```
[1] 1.722933
```

(b) The 5th percentile (0.95 critical value $t_{0.95}(4)$) is
$t_{0.95}(4) =$ (i) $\mathbf{-2.13^o}$ (ii) $\mathbf{-2.01^o}$ (iii) $\mathbf{-1.23^o}$ (iv) $\mathbf{-1.02^o}$

```
qt(0.05,4) # Student-t, 4 df, 5th percentile

[1] -2.131847
```

or, using Table C.2
$t_{0.05}(4) =$ (i) $\mathbf{-2.13^o}$ (ii) $\mathbf{-2.01^o}$ (iii) $\mathbf{-1.23^o}$ (iv) $\mathbf{-1.02^o}$
Make number found in Table C.2 negative because Student-$t$ is symmetric around zero and 5th percentile is below zero.

(c) The 58th percentile (0.42 critical value $t_{0.42}(4)$) is (i) $\mathbf{0.12^o}$ (ii) $\mathbf{0.16^o}$ (iii) $\mathbf{0.18^o}$ (iv) $\mathbf{0.22^o}$

```
qt(0.58,4) # Student-t, 4 df, 58th percentile

[1] 0.2153901
```

(d) The 23rd percentile (0.77 critical value $t_{0.77}(4)$) is (i) $\mathbf{-2.01^o}$ (ii) $\mathbf{-1.32^o}$ (iii) $\mathbf{-0.82^o}$ (iv) $\mathbf{-0.56^o}$

```
qt(0.23,4) # Student-t, 4 df, 23rd percentile

[1] -0.8165961
```

7. *F distribution.*

(a) assuming $(n, d) = (2, 2)$ degrees of freedom
$P(F < 3.9) =$ (i) **0.80** (ii) **0.84** (iii) **0.89** (iv) **0.92**.

```
pf(3.9,2,2) # F, 2,2 df, P(F < 3.9)
```

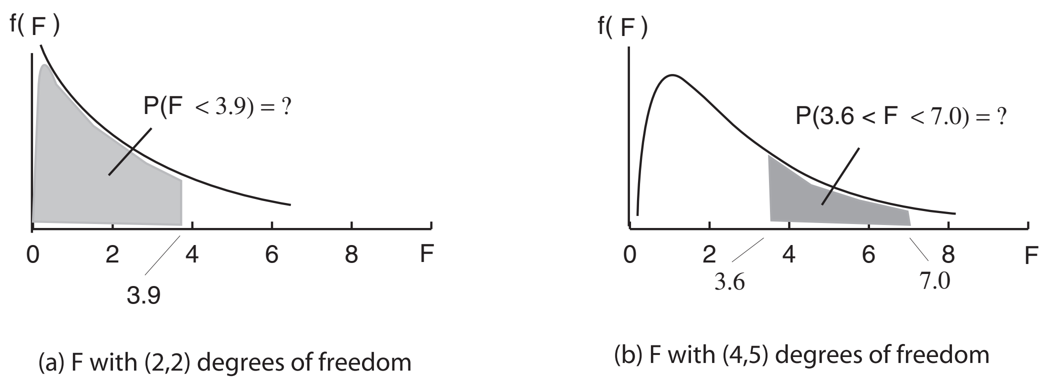



Figure 3.24: $F$ distributions

```
[1] 0.7959184
```

(b) assuming $(n, d) = (4, 5)$ degrees of freedom
$P(3.6 < F < 7) =$ (i) **0.05** (ii) **0.07** (iii) **0.09** (iv) **0.11**.

```
pf(7,4,5) - pf(3.6,4,5) # F, 4,5 df, P(3.6 < T < 7)
[1] 0.06840958
```

(c) assuming $(n, d) = (4, 5)$ degrees of freedom

$$\mu = \frac{d}{d-2} \approx$$

(i) **1.67** (ii) **1.73** (iii) **1.75** (iv) **1.79**.

(d) assuming $(n, d) = (4, 5)$ degrees of freedom

$$\sigma = \sqrt{\frac{2d^2(n+d-2)}{n(d-2)^2(d-4)}} = \sqrt{\frac{2(5)^2(4+5-2)}{4(5-2)^2(5-4)}} \approx$$

(i) **2.71** (ii) **3.12** (iii) **3.75** (iv) **4.79**.

8. *Critical value $f_p(n, d)$ (percentile).*

(a) The 0.05 critical value (95th percentile), $(n, d) = (2, 2)$ df
$f_{0.05}(2, 2) =$ (i) **16** (ii) **17** (iii) **18** (iv) **19**,

```
qf(0.95,2,2) # F, 2,2 df, 0.05 critical value, 95th percentile
[1] 19
```

or, using Table C.4
$f_{0.95}(4, 5) =$ (i) **16** (ii) **17** (iii) **18** (iv) **19**.

Table C.4 critical values restricted to select $(n, d)$.

(b) The 0.95 critical value (5th percentile), $(n, d) = (4, 5)$ df
$f_{0.95}(4, 5) =$ (i) **0.11** (ii) **0.16** (iii) **0.23** (iv) **0.34**.

```
qf(0.05,4,5) # F, 4,5 df, 0.95 critical value, 5th percentile
[1] 0.1598451
```

## 3.10 Approximating the Binomial Distribution

*Limit Theorem of De Moivre and Laplace* says let $X$ be $b(n, p)$, then, as $n \to \infty$,

$$X \text{ approaches } N(np, np(1-p)),$$

and this is typically a close approximation if $np \geq 5$, $n(1-p) \geq 5$; furthermore, the approximation improves if a 0.5 *continuity correction* is used,

$$P(a \leq X \leq b) \approx P(a - 0.5 \leq Y \leq b + 0.5)$$

where $Y$ is $N(np, np(1-p))$. A related result is $\frac{X}{n}$ approaches $N(p, \frac{p(1-p)}{n})$. Both results are related to the Central Limit Theorem.

*Bernoulli's Law of Large Numbers* says let random variable $X$ be number of successes (of observing event $A$) in $n$ trials and $p = P(A)$, then for any small $\epsilon > 0$,

$$P\left(\left|\frac{X}{n} - p\right| \leq \epsilon\right) \to 1, \quad \text{as } n \to \infty.$$

This essentially says the sample proportion, $\frac{X}{n}$, tends to the population proportion, $p$ for large $n$. De Moivre and Laplace's limit theorem is "stronger" in the sense it says not only $\frac{X}{n}$ converges to $p$ but also "how" it converges, it converges to a normal distribution "around" $p$, but "weaker" in the sense $X$ must be $b(n, p)$ whereas Bernoulli's Law does not require a "starting" distribution.

**Exercise 3.10 (Approximating the Binomial Distribution)**
A lawyer estimates she has a 40% ($p = 0.4$) of winning each of her next 10 ($n = 10$) cases. Assume number of wins $X$ is $b(10, 0.4)$.

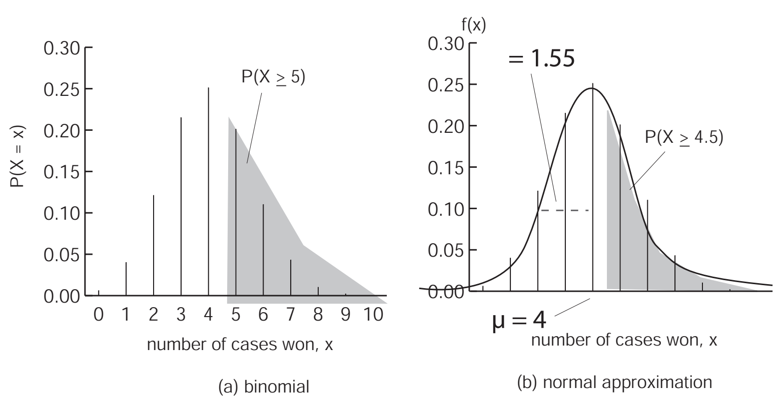



Figure 3.25: Normal approximation to Binomial

1. *Check conditions for approximation.* Since

$$np = 10(0.4) = 4 < 5, \quad n(1-p) = 10(0.6) = 6 > 5,$$

   conditions for approximation (i) **have** (ii) **have not** been satisfied (but they are close enough to continue).

2. Chance she has at least 5 wins

   (a) *(exact) binomial*
   $P(X \geq 5) = 1 - P(X \leq 4) =$
   (i) **0.37** (ii) **0.42** (ii) **0.57**.

```
1 - pbinom(4,10,0.4) # binomial(10,0.4), P(X ge 5) = 1 - P(X le 4)
```

```
[1] 0.3668967
```

   (b) *(approximate) normal*
   since $\mu = np = 10(0.4) =$
   (i) **4** (ii) **5** (iii) **6**
   and $\sigma = \sqrt{np(1-p)} = \sqrt{10(0.4)(1-0.4)} = \sqrt{10(0.4)(0.6)} \approx$
   (i) **1.2322345** (ii) **1.5491934** (iii) **1.9943345**,
   then
   $P(X \geq 5) \approx P(X \geq 4.5) = 1 - P(X \leq 4.5) \approx$
   **0.37** (ii) **0.42** (ii) **0.57**.

```
1 - pnorm(4.5,10*0.4,sqrt(10*0.4*0.6)) # N(mu,sigma), mu = np, sigma = sqrt(np(1-p)), P(X ge 4.5)
```

```
[1] 0.3734428
```

3. Chance she has *more than* 5 wins is

   (a) *(exact) binomial*
   $P(X > 5) = P(X \geq 6) = 1 - P(X \leq 5) =$ (i) **0.17** (ii) **0.21** (iii) **0.24**.

```
1 - pbinom(5,10,0.4) # binomial(10,0.4), P(X > 5) = 1 - P(X le 5)
```

```
[1] 0.1662386
```

   (b) *(approximate) normal*
   $P(X > 5) = P(X \geq 6) \approx P(X \geq 5.5) = 1 - P(X \leq 5.5) =$
   (i) **0.17** (ii) **0.21** (iii) **0.24**.

```
1 - pnorm(5.5,10*0.4,sqrt(10*0.4*0.6)) # P(X ge 5.5)
```

```
[1] 0.1664608
```

4. Chance she has at most 5 wins is,

   (a) *(exact) binomial*
   $P(X \leq 5) =$
   (i) **0.83** (ii) **0.92** (iii) **0.99**.

```
pbinom(5,10,0.4) # binomial(10,0.4), P(X le 5)
```

```
[1] 0.8337614
```

(b) *(approximate) normal:*
$P(X \leq 5) \approx P(X \leq 5.5) \approx$
(i) **0.83** (ii) **0.92** (iii) **0.99**.

```
pnorm(5.5,10*0.4,sqrt(10*0.4*0.6)) # P(X ge 5.5)
```

```
[1] 0.8335392
```

5. Chance she has *exactly* 5 wins is,

(a) *(exact) binomial*
$P(X = 5) =$ (i) **0.17** (ii) **0.20** (iii) **0.24**.

```
dbinom(5,10,0.4) # binomial(10,0.4), P(X = 5)
```

```
[1] 0.2006581
```

(b) *(approximate) normal*
$P(X = 5) \approx P(4.5 \leq X \leq 5.5) = P(X \leq 5.5) - P(X \leq 4.5) \approx$
(i) **0.17** (ii) **0.21** (iii) **0.24**.

```
pnorm(5.5,10*0.4,sqrt(10*0.4*0.6)) - pnorm(4.5,10*0.4,sqrt(10*0.4*0.6))  # P(4.5 le X le 5.5)
```

```
[1] 0.206982
```

| CONTINUOUS | $f(x)$ | $M(t)$ | $\mu$ | $\sigma^2$ |
|---|---|---|---|---|
| Uniform | $\frac{1}{b-a}$ | $\frac{e^{tb}-e^{ta}}{t(b-a)}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| Exponential | $\lambda e^{-\lambda x}$ | $\frac{\lambda}{\lambda-t}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Normal | $\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$ | $\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$ | $\mu$ | $\sigma^2$ |
| Gamma | $\frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$ | $\left(1 - \frac{t}{\lambda}\right)^{-r}$ | $\frac{r}{\lambda}$ | $\frac{r}{\lambda^2}$ |
| Chi-square | $\frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$ | $(1-2t)^{-\frac{n}{2}}$ | $n$ | $2n$ |
| Student-t | $\frac{\Gamma\left[\frac{n+1}{2}\right]}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ | undefined | 0 | $\frac{n}{n-2}$ |
| F | $\frac{\Gamma\left[\frac{n+1}{2}\right] n^{\frac{n}{2}} d^{\frac{d}{2}} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)(d+nx)^{\frac{n+d}{2}}}$ | undefined | $\frac{d}{d-2}$ | $\frac{2d^2(n+d-2)}{n(d-2)^2(d-4)}$ |