

Required Problems

1. Let
- Y
- be a continuous random variable with PDF

$$f_Y(y) = \begin{cases} (3/2)y^2 + y & \text{if } 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}$$

- (a) Find the mean of
- Y
- .

$$\mathbb{E}[Y] = \int_0^1 y \left[\frac{3}{2}y^2 + y \right] dy \quad (\text{by def. of expected value})$$

$$= \int_0^1 \frac{3}{2}y^3 + y^2 \quad (\text{simplifying})$$

$$= \left[\frac{3}{8}y^4 + \frac{1}{3}y^3 \right]_0^1 \quad (\text{taking the integral})$$

$$\mathbb{E}[Y] = \frac{17}{24} \quad (\text{evaluating})$$

- (b) Find the variance of
- Y
- .

$$\mathbb{E}[Y^2] = \int_0^1 y^2 \left[\frac{3}{2}y^2 + y \right] dy \quad (\text{by def. of expected value})$$

$$= \int_0^1 \frac{3}{2}y^4 + y^3 dy \quad (\text{simplifying})$$

$$= \left[\frac{3}{10}y^5 + \frac{1}{4}y^4 \right]_0^1 \quad (\text{taking the integral})$$

$$\mathbb{E}[Y^2] = \frac{11}{20} \quad (\text{evaluating})$$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad (\text{by def. of variance})$$

$$\text{Var}(X) = \frac{11}{20} - \left[\frac{17}{24} \right]^2 \quad (\text{plugging in values})$$

$$\text{Var}(X) = \frac{139}{2880} \approx 0.0483 \quad (\text{simplifying})$$

2. Let
- Y
- be a random variable with probability density function given by

$$f_Y(y) = 2(1 - y), \quad y \in [0, 1]$$

- (a) Find the PDF of
- $U = 2Y - 1$
- .

Because this is a one-to-one transformation, the transformation theorem is applicable. The problem can still be solved, however, by employing the distribution function method:

$$F_U(u) = \mathbb{P}(U \leq u) \quad (\text{by def. of the CDF})$$

$$= \mathbb{P}(2Y - 1 \leq u) \quad (\text{by def. of } U)$$

$$= \mathbb{P}\left(Y \leq \frac{u+1}{2}\right) \quad (\text{rearranging})$$

$$= F_Y\left(\frac{u+1}{2}\right) \quad (\text{by def. of the CDF})$$

Differentiating to find the PDF:

$$\begin{aligned}
 f_U(u) &= f_Y\left(\frac{u+1}{2}\right) \left(\frac{1}{2}\right) && \text{(differentiating w.r.t. } u) \\
 &= 2 \left(1 - \frac{u+1}{2}\right) \left(\frac{1}{2}\right) && \text{(plugging into the PDF of } Y) \\
 f_U(u) &= \begin{cases} \frac{1}{2} - \frac{u}{2} & \text{if } u \in [-1, 1] \\ 0 & \text{else} \end{cases} && \text{(simplifying)}
 \end{aligned}$$

Alternatively, the problem may be solved via the transformation method:

$$\begin{aligned}
 Y &= \frac{U}{2} + \frac{1}{2} && \text{(solving for } Y) \\
 \frac{dY}{dU} &= \frac{1}{2} && \text{(differentiating)} \\
 f_U(u) &= 2 \left(1 - \left[\frac{U}{2} + \frac{1}{2}\right]\right) \left|\frac{1}{2}\right| && \text{(by the trans. meth.)}
 \end{aligned}$$

Again, noting the change of support: $u \in [-1, 1]$:

$$f_U(u) = \begin{cases} \frac{1}{2} - \frac{u}{2} & \text{if } u \in [-1, 1] \\ 0 & \text{else} \end{cases} \quad \text{(the formal PDF)}$$

(b) **Find the PDF of $W = 1 - 2Y$.**

Again, this is a one-to-one transformation:

$$\begin{aligned}
 Y &= \frac{1}{2} - \frac{W}{2} && \text{(solving for } Y) \\
 \frac{dY}{dW} &= -\frac{1}{2} && \text{(differentiating)} \\
 f_W(w) &= 2 \left(1 - \left[\frac{1}{2} - \frac{W}{2}\right]\right) \left|-\frac{1}{2}\right| && \text{(by the trans. meth.)}
 \end{aligned}$$

Again, we need to note the change of support: $w \in [-1, 1]$:

$$f_U(u) = \begin{cases} \frac{1}{2} + \frac{W}{2} & \text{if } w \in [-1, 1] \\ 0 & \text{else} \end{cases} \quad \text{(the formal PDF)}$$

(c) **Find the PDF of $Z = Y^2$.**

This is not a one-to-one function generally; however, on the interval $[0, 1]$, y^2 is one-to-one. Thus, the transformation method can be used again:

$$\begin{aligned}
 Y &= \sqrt{Z} && \text{(solving for } Y) \\
 \frac{dY}{dZ} &= \frac{1}{2\sqrt{Z}} && \text{(differentiating)} \\
 f_Z(z) &= 2 \left(1 - \sqrt{z}\right) \left|\frac{1}{2\sqrt{z}}\right| && \text{(by the trans. meth.)}
 \end{aligned}$$

Again, we need to note the “change” of support: $z \in (0, 1]$:

$$f_Z(z) = \begin{cases} \frac{1}{\sqrt{z}} - 1 & \text{if } z \in (0, 1] \\ 0 & \text{else} \end{cases} \quad (\text{the formal PDF})$$

Note that we must have $z \in (0, 1]$ (the lower bound is non-inclusive), since the function will be undefined if $z = 0$. That said, it's a continuous random variable, so the probability that $y = 0$ is zero.

3. Consider the multivariate distribution characterized by the PDF

$$f_{XY} = 6(1 - y), \quad 0 \leq x \leq y \leq 1$$

(a) Find the conditional expectation $\mathbb{E}[X|Y = y]$.

The conditional expectation requires the conditional distribution of X given Y . The first step is to find the marginal distribution of Y :

$$f_Y(y) = \int_0^y 6(1 - y) \, dx \quad (\text{integrating over } x)$$

$$= 6(1 - y)x \Big|_0^y \quad (\text{integrating})$$

$$f_Y(y) = 6y(1 - y) \quad (\text{simplifying})$$

where $y \in [0, 1]$. Next, the conditional distribution is given by:

$$f(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad (\text{by def. of the cond. PDF})$$

$$= \frac{6(1 - y)}{6y(1 - y)} \quad (\text{plugging in functions})$$

$$f(x|y) = \frac{1}{y} \quad (\text{simplifying})$$

where $x \in [0, y]$. To find the conditional expectation, integrate over the conditional PDF:

$$\mathbb{E}[X|Y = y] = \int_0^y x \left(\frac{1}{y} \right) \, dx \quad (\text{finding the exp. value})$$

$$= \frac{x^2}{2y} \Big|_0^y \quad (\text{taking the integral})$$

$$\mathbb{E}[X|Y = y] = \frac{y}{2} \quad (\text{evaluating})$$

(b) Find the covariance of X and Y .

To find the covariance, three values are required: $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$. The conditional PDF of y was found in part (a); the marginal PDFs of X is given by:

$$f_X(x) = \int_x^1 6(1 - y) \, dy \quad (\text{integrating over } y)$$

$$= 6y - 3y^2 \Big|_x^1 \quad (\text{integrating})$$

$$f_X(x) = 3x^2 - 6x + 3 \quad (\text{simplifying})$$

where $x \in [0, 1]$. Using the conditional PDFs, the expected values are:

$$\mathbb{E}[X] = \int_0^1 x(3x^2 - 6x + 3)dx \quad (\text{finding the exp. value})$$

$$= \frac{3}{4}x^4 - 2x^3 + \frac{3}{2}x^2 \Big|_0^1 \quad (\text{integrating})$$

$$\mathbb{E}[X] = \frac{1}{4} \quad (\text{evaluating})$$

$$\mathbb{E}[Y] = \int_0^1 y(6y - 6y^2) dy \quad (\text{finding the exp. value})$$

$$= 2y^3 - \frac{6}{4}y^4 \Big|_0^1 \quad (\text{integrating})$$

$$\mathbb{E}[Y] = \frac{1}{2} \quad (\text{evaluating})$$

$$\mathbb{E}[XY] = \int_0^1 \int_0^y xy6(1-y) dx dy \quad (\text{finding the exp. value})$$

$$= \int_0^1 \left(\left[3x^2y - 3x^2y^2 \right]_0^y \right) dy \quad (\text{taking the inner integral})$$

$$= \int_0^1 3y^3 - 3y^4 dy \quad (\text{evaluating})$$

$$= \left[\frac{3}{4}y^4 - \frac{3}{5}y^5 \right]_0^1 \quad (\text{taking the outer integral})$$

$$\mathbb{E}[XY] = \frac{3}{20} \quad (\text{evaluating})$$

Putting the pieces together to find the covariance:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (\text{by def. of the covariance})$$

$$= \frac{3}{20} - \left(\frac{1}{4} \right) \left(\frac{1}{2} \right) \quad (\text{plugging in values})$$

$$\text{Cov}(X, Y) = \frac{1}{40} \quad (\text{simplifying})$$

4. Let X_1, \dots, X_n be a random sample from a distribution with PMF

$$f(x_i|\theta) = \begin{cases} \theta(1-\theta)^{x_i-1} & \text{if } x = 1, 2, 3 \dots \\ 0 & \text{else} \end{cases}$$

where $\theta \in (0, 1)$.

(a) Find the method of moments estimator for θ .

Finding the first population momen $\mathbb{E}[X_i]$:

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x\theta(1-\theta)^{x-1} \quad (\text{by def. of the expected value})$$

Note that this is a little bit more complicated than a typical geometric series; writing out the first few terms of the sum:

$$\mathbb{E}[X] = \theta [1 + 2(1 - \theta) + 3(1 - \theta)^2 + 4(1 - \theta)^3 + \dots] \quad (\text{writing out terms})$$

$$(1 - \theta)\mathbb{E}[X] = \theta [(1 - \theta) + 2(1 - \theta)^2 + 3(1 - \theta)^3 + \dots] \quad (\text{multiplying by } (1 - \theta))$$

$$\mathbb{E}[X] - (1 - \theta)\mathbb{E}[X] = \theta [1 + (1 - \theta) + (1 - \theta)^2 + (1 - \theta)^3 + \dots] \quad (\text{subtracting the two})$$

The right-hand-side is now a geometric series with $(1 - \theta) < 1$:

$$\mathbb{E}[X] - (1 - \theta)\mathbb{E}[X] = \theta \left(\frac{1}{1 - (1 - \theta)} \right) \quad (\text{the sum of a geometric series})$$

$$\mathbb{E}[X] = \frac{1}{\theta} \quad (\text{solving for } \mathbb{E}[X])$$

This is the first population moment. Recall that the first sample moment is given by:

$$m_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{the first sample moment})$$

Using the population and sample moments, we can find the MME:

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{\hat{\theta}} \quad (\text{matching moments})$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n X_i} \quad (\text{solving for } \hat{\theta})$$

$$\hat{\theta}_{mme} = \frac{1}{\bar{X}} \quad (\text{simplifying})$$

(b) **Find the maximum likelihood estimator for θ .**

Because we have a random sample, the likelihood function is given by:

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i|\theta) \quad (\text{the likelihood})$$

$$= \prod_{i=1}^n \theta(1 - \theta)^{x_i - 1} \quad (\text{plugging in the PDF})$$

$$\mathcal{L}(\theta|\mathbf{x}) = \theta^n (1 - \theta)^{\sum_{i=1}^n x_i - n} \quad (\text{multiplying})$$

Taking a logarithmic transformation to make it easier to work with:

$$\ln(\mathcal{L}(\cdot)) = n \ln(\theta) + \left(\sum_{i=1}^n x_i - n \right) \ln(1 - \theta) \quad (\text{taking the log})$$

$$\frac{\partial \ln(\mathcal{L}(\cdot))}{\partial \theta} = \frac{n}{\hat{\theta}} - \frac{\sum_{i=1}^n x_i - n}{1 - \hat{\theta}} = 0 \quad (\text{the FOC})$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i} \quad (\text{solving for } \hat{\theta})$$

$$\hat{\theta}_{mle} = \frac{1}{\bar{X}} \quad (\text{simplifying})$$

Practice Problems

5. Let X be a discrete random variable with PMF $f_X(x)$, given in the following table.

x	1	2	3	4
$f_X(x)$	0.4	0.3	0.2	0.1

Find the following:

(a) $\mathbb{E}[X]$

$$\mathbb{E}[X] = 1(0.4) + 2(0.3) + 3(0.2) + 4(0.1) \quad (\text{by def. of expected value})$$

$$\mathbb{E}[X] = 2 \quad (\text{simplifying})$$

(b) $\mathbb{E}[1/X]$

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{1}(0.4) + \frac{1}{2}(0.3) + \frac{1}{3}(0.2) + \frac{1}{4}(0.1) \quad (\text{by def. of expected value})$$

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{77}{120} \approx 0.642 \quad (\text{simplifying})$$

(c) $\mathbb{E}[X^2 - 1]$

$$\mathbb{E}[X^2 - 1] = [1^2 - 1](0.4) + [2^2 - 1](0.3) + [3^2 - 1](0.2) + [4^2 - 1](0.1) \quad (\text{by def. of expected value})$$

$$\mathbb{E}[X^2 - 1] = 4 \quad (\text{simplifying})$$

(d) $\text{Var}[X]$

$$\mathbb{E}[X^2] = 1^2(0.4) + 2^2(0.3) + 3^2(0.2) + 4^2(0.1) \quad (\text{by def. of expected value})$$

$$\mathbb{E}[X^2] = 5 \quad (\text{calculating})$$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad (\text{by def. of variance})$$

$$\text{Var}(X) = 5 - [2]^2 \quad (\text{plugging in values})$$

$$\text{Var}(X) = 1 \quad (\text{simplifying})$$

6. A binomially-distributed random variable X has the PMF

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \dots, n\}$$

where $0 < p < 1$.

(a) Show that the expected value of X is np .

$$\mathbb{E}[X] = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{by def. of expected value})$$

$$= 0 \binom{n}{0} p^0 (1-p)^{n-0} + \sum_{x=1}^n x \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{breaking up the sum})$$

$$= \sum_{x=1}^n x \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{simplifying})$$

Now, employing the formulas from combinatorics:

$$\mathbb{E}[X] = \sum_{x=1}^n x \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{by the comb. formula})$$

$$= np \sum_{x=1}^n x \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \quad (\text{factoring out an } np)$$

Note that we can simplify the factorial formula slightly using $\frac{x}{x!} = \frac{1}{(x-1)!}$:

$$= np \sum_{x=1}^n \left(\frac{(n-1)!}{(x-1)!(n-x)!} \right) p^{x-1} (1-p)^{n-x} \quad (\text{canceling an } x)$$

Also note that $n-x = (n-1) - (x-1)$, which lets us again modify the equation:

$$= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!([n-1] - [x-1])!} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

With this last change, we can simplify the factorial expression back into our combinatorics notation:

$$= np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)} \quad (\text{by the comb. formula})$$

Now the expression in the sum is simply the PMF of a binomial distribution, with $n-1$ observations and $x-1$ successes. Further, because it's summed over its whole support:

$$= np[1] \quad (\text{by def. of a PMF})$$

$$\mathbb{E}[X] = np \quad (\text{simplifying})$$

(b) **Show that the variance of X is $np(1-p)$.**

$$\mathbb{E}[X^2] = \sum_{x=0}^n x^2 \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{by def. of expected value})$$

$$= np \sum_{x=1}^n x \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)} \quad (\text{from part (a)})$$

Now let's define new values, $y = x-1$ and $m = n-1$. Then we can rewrite the sum:

$$= np \sum_{y=0}^m (y+1) \binom{m}{y} p^y (1-p)^{m-y} \quad (\text{plugging in for } x-1, n-1)$$

We can break up the sum across $(y+1)$:

$$= np \left[\sum_{y=0}^m y \binom{m}{y} p^y (1-p)^{m-y} + \sum_{y=0}^m \binom{m}{y} p^y (1-p)^{m-y} \right]$$

Note now that the first sum is simply $E[Y]$, while the second is the sum of the PDF over its support:

$$= np[mp + 1] \quad (\text{from part (a)})$$

$$= np[(n-1)p + 1] \quad (\text{plugging in for } m)$$

$$\mathbb{E}[X^2] = (np)^2 + np(1-p) \quad (\text{rearranging})$$

Now we can simply plug values into our variance formula to find the variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad (\text{by def. of variance})$$

$$= (np)^2 + np(1-p) - (np)^2 \quad (\text{plugging in values})$$

$$\text{Var}(X) = np(1-p)$$

7. Let Y be a random variable with mean μ and variance σ^2 . If a and b are constants, show that:

(a) $\mathbb{E}[aY + b] = a\mu + b$

For a continuous random variable:

$$\mathbb{E}[aY + b] = \int_{-\infty}^{\infty} (ay + b) f_Y(y) dy \quad (\text{by def. of expected value})$$

$$= \int_{-\infty}^{\infty} ay f_Y(y) dy + \int_{-\infty}^{\infty} b f_Y(y) dy \quad (\text{breaking up the integral})$$

$$= a \int_{-\infty}^{\infty} y f_Y(y) dy + b \int_{-\infty}^{\infty} f_Y(y) dy \quad (\text{pulling out constants})$$

Note that the second integral is over the entire real line (including the support of $f_Y(y)$). Thus:

$$\mathbb{E}[aY + b] = a \mathbb{E}[Y] + b \quad (\text{by def. of expected value})$$

The argument is analogous for discrete random variables, using sums instead of integrals.

(b) $\text{Var}(aY + b) = a^2\sigma^2$

$$\text{Var}(aY + b) = \mathbb{E}[(aY + b)^2] - \mathbb{E}[aY + b]^2 \quad (\text{by def. of variance})$$

$$= \mathbb{E}[a^2Y^2 + 2abY + b^2] - \mathbb{E}[aY + b]^2 \quad (\text{expanding})$$

$$= a^2\mathbb{E}[Y^2] + 2ab\mathbb{E}[Y] + b^2 - (a\mathbb{E}[Y] + b)^2 \quad (\mathbb{E} \text{ is a linear operator})$$

$$= a^2\mathbb{E}[Y^2] + 2ab\mathbb{E}[Y] + b^2 - (a^2\mathbb{E}[Y]^2 + 2ab\mathbb{E}[Y] + b^2) \quad (\text{expanding})$$

$$= a^2\mathbb{E}[Y^2] - a^2\mathbb{E}[Y]^2 \quad (\text{simplifying})$$

$$\text{Var}(aY + b) = a^2\text{Var}(Y) \quad (\text{by def. of variance})$$

8. Let X be a standard normal random variable with PDF

$$f_X(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \quad x \in (-\infty, \infty)$$

- (a) **Show that the MGF of X is $M_X(t) = \exp\{(1/2)t^2\}$ (hint: complete the square in the exponent).**

$$\begin{aligned}
 M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx && \text{(by def. of the MGF)} \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{tx - \frac{x^2}{2}\right\} dx && \text{(simplifying)} \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(-2tx + x^2)\right\} dx && \text{(rearranging)} \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(t^2 - 2tx + x^2 - t^2)\right\} dx && \text{(completing the square)} \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(t^2 - 2tx + x^2) + \frac{1}{2}t^2\right\} dx && \text{(pulling out a term)} \\
 &= \int_{-\infty}^{\infty} \exp\left\{\frac{1}{2}t^2\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(t^2 - 2tx + x^2)\right\} dx && \text{(splitting up the exponent)} \\
 &= \exp\left\{\frac{1}{2}t^2\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x - t)^2\right\} dx && \text{(pulling out a constant, rearranging)}
 \end{aligned}$$

Note that the integral is now over entire support of the PDF of a $N(t, 1)$ random variable (a fairly common trick in probability and statistics). Thus

$$M_X(t) = \exp\left\{\frac{1}{2}t^2\right\}$$

- (b) **Use the MGF of X to calculate $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$.**

We already know that $\mu = 0$ and $\sigma^2 = 1$, but for the sake of practice:

$$\begin{aligned}
 \frac{d}{dt} M_X(t) &= t \exp\left\{\frac{1}{2}t^2\right\} && \text{(differentiating w.r.t. } t) \\
 \frac{d}{dt} M_X(t)|_{t=0} &= 0 && \text{(evaluating at zero)} \\
 \mathbb{E}[X] &= 0 && \text{(by properties of MGFs)} \\
 \frac{d^2}{dt^2} M_X(t) &= \exp\left\{\frac{1}{2}t^2\right\} + t^2 \exp\left\{\frac{1}{2}t^2\right\} && \text{(the second derivative)} \\
 \frac{d^2}{dt^2} M_X(t)|_{t=0} &= 1 && \text{(evaluating at zero)} \\
 \mathbb{E}[X^2] &= 1 && \text{(by properties of MGFs)}
 \end{aligned}$$

- (c) **Use the MGF to show that the sum of two independent standard normal random variables is distributed normally with $\mu = 0$ and $\sigma^2 = 2$.**

Let Z_1 and Z_2 be independent standard normal random variables. Let $Y = Z_1 + Z_2$.

$$M_{Z_i}(t) = \exp\left\{\frac{1}{2}t^2\right\} \quad \text{(by part (a))}$$

Recall the theorem regarding linear combinations of independent random variables and MGFs:

$$\begin{aligned}
 M_Y(t) &= M_{Z_1}(t)M_{Z_2}(t) && \text{(Y is a linear combo. of } Z_1, Z_2\text{)} \\
 &= \exp\left\{\frac{1}{2}t\right\}\exp\left\{\frac{1}{2}t\right\} && \text{(plugging in MGFs)} \\
 &= \exp\left\{\frac{1}{2}t + \frac{1}{2}t\right\} && \text{(summing exponents)} \\
 M_Y(t) &= \exp\left\{0t + \frac{1}{2}(2)t\right\} && \text{(rearranging)}
 \end{aligned}$$

This takes the form: $\exp\{\mu t + (1/2)\sigma^2 t\}$. Because MGFs uniquely identify distributions, we know that Y must be a normal random variable. Further, in the MGF of Y , it is clear that $\mu = 0$ and $\sigma^2 = 2$. Thus, $Y \sim N(0, 2)$.

9. For each of the following random variables, find the PDF of Y .

(a) $f_X(x) = \frac{1}{2}(1+x)$, $x \in (-1, 1)$ **and** $Y = X^2$

This is NOT a one-to-one transformation, so we can't use the transformation theorem. Instead, using the CDF approach:

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) && \text{(by def. of the CDF)} \\
 &= P(X^2 \leq y) && \text{(plugging in for } Y\text{)} \\
 &= P(-\sqrt{y} \leq X \leq \sqrt{y}) && \text{(isolating } X\text{)} \\
 F_Y(y) &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) && \text{(by properties of the CDF)}
 \end{aligned}$$

We can differentiate to find an expression for the PDF:

$$\begin{aligned}
 f_Y(y) &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} && \text{(differentiating w.r.t. } y\text{)} \\
 &= \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] && \text{(simplifying)} \\
 &= \frac{1}{2\sqrt{y}} \left[\frac{1}{2}(1+\sqrt{y}) + \frac{1}{2}(1-\sqrt{y}) \right] && \text{(plugging in for } f_X\text{)} \\
 &= \frac{1}{2\sqrt{y}} && \text{(simplifying)}
 \end{aligned}$$

Accounting for the change in support: $y \in (0, 1)$:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{if } y \in (0, 1) \\ 0 & \text{else} \end{cases} \quad \text{(the formal PDF)}$$

(b) $f_X(x) = 60x^3(1-x)^2$, $x \in (0, 1)$ **and** $Y = \log(X)$

This is a one-to-one transformation, so again, we can use our handy transformation theorem:

$$\begin{aligned}
 Y &= \log[X] && \text{(the transformed RV)} \\
 X &= e^Y && \text{(\log[X] is monotone)}
 \end{aligned}$$

$$\frac{dX}{dY} = e^Y \quad (\text{differentiating})$$

$$f_Y(y) = 60(e^y)^3(1 - e^y)^2 |e^y| \quad (\text{by the trans. method})$$

$$f_Y(y) = \begin{cases} 60e^{4y}(1 - e^y)^2 & \text{if } y \in (-\infty, 0) \\ 0 & \text{else} \end{cases} \quad (\text{the formal PDF})$$

(c) $f_X(x) = (1+x)^2/9$, $x \in (-1, 2)$ **and** $Y = (X-1)^2$

Consider the first case: $-1 < x < 0$ (i.e. the values of x that map one-to-one to y):

$$F_Y(y) = P[Y \leq y] \quad (\text{by def. of the CDF})$$

$$= P[(X-1)^2 \leq y] \quad (\text{plugging in for } Y)$$

$$= P[-\sqrt{y} \leq X-1] \quad (\text{rearranging the prob.})$$

$$= P[1 - \sqrt{y} \leq X] \quad (\text{isolating } X)$$

$$= 1 - P[X \leq 1 - \sqrt{y}] \quad (\text{by properties of complements})$$

$$F_Y(y) = 1 - F_X(1 - \sqrt{y}) \quad (\text{by def. of the CDF})$$

$$f_Y(y) = -f_X(1 - \sqrt{y}) \cdot \left(-\frac{1}{2\sqrt{y}}\right) \quad (\text{differentiating w.r.t. } y)$$

$$= \frac{(1 + [1 - \sqrt{y}])^2}{9} \cdot \frac{1}{2\sqrt{y}} \quad (\text{plugging in the function } f(x))$$

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left(\frac{(2 - \sqrt{y})^2}{9} \right), \quad y \in (1, 4) \quad (\text{simplifying})$$

Note that this is only part of the PDF, i.e., the part defined over $x \in (-1, 0)$. Consider the second case: $0 < x < 2$ (i.e., the values of x that do not map one-to-one to y):

$$F_Y(y) = P[Y \leq y] \quad (\text{by def. of the CDF})$$

$$= P[-\sqrt{y} \leq X-1 \leq \sqrt{y}] \quad (\text{rearranging the prob.})$$

$$= P[X \leq 1 + \sqrt{y}] - P[X \leq 1 - \sqrt{y}] \quad (\text{isolating } X)$$

$$F_Y(y) = F_X(1 + \sqrt{y}) - F_X(1 - \sqrt{y}) \quad (\text{by def. of the CDF})$$

$$f_Y(y) = f_X(1 + \sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(1 - \sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \quad (\text{differentiating w.r.t. } y)$$

$$= \frac{(1 + [1 + \sqrt{y}])^2}{9} \cdot \frac{1}{2\sqrt{y}} + \frac{(1 + [1 - \sqrt{y}])^2}{9} \cdot \frac{1}{2\sqrt{y}} \quad (\text{plugging in the function } f(x))$$

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left(\frac{8 + 2y}{9} \right), \quad y \in (0, 1) \quad (\text{simplifying})$$

Putting the two pieces together:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left(\frac{8 + 2y}{9} \right) & \text{if } 0 < y < 1 \\ \frac{1}{2\sqrt{y}} \left(\frac{(2 - \sqrt{y})^2}{9} \right) & \text{if } 1 < y < 4 \\ 0 & (\text{else}) \end{cases} \quad (\text{the formal PDF})$$

10. Consider the joint distribution of Y_1 and Y_2 given by the following table:

		x_1	
		0	1
x_2	0	0.38	0.17
	1	0.14	0.02
	2	0.24	0.05

(a) Find the marginal PMFs for X_1 and X_2 .

To find the marginal PMFs, we sum over the other variable:

$$f_{X_1}(x_1) = \begin{cases} 0.76 & \text{if } x_1 = 0 \\ 0.24 & \text{if } x_1 = 1 \\ 0 & \text{else} \end{cases} \quad f_{X_2}(x_2) = \begin{cases} 0.55 & \text{if } x_2 = 0 \\ 0.16 & \text{if } x_2 = 1 \\ 0.29 & \text{if } x_2 = 2 \\ 0 & \text{else} \end{cases}$$

(b) Find the conditional PMF for X_2 given $X_1 = 0$.

Recall that the marginal PMF is given by $f(x_2|X_1 = x_1) = f(x_1, x_2)/f(x_1)$. From (a), we know that $f_{X_1}(0) = 0.76$. Thus:

$$f(x_2|X_1 = 0) = \begin{cases} \frac{0.38}{0.76} & \text{if } x_2 = 0 \\ \frac{0.14}{0.76} & \text{if } x_2 = 1 \\ \frac{0.24}{0.76} & \text{if } x_2 = 2 \end{cases} \quad (\text{take the ratio})$$

$$f(x_2|X_1 = 0) = \begin{cases} \frac{1}{2} & \text{if } x_2 = 0 \\ \frac{7}{38} & \text{if } x_2 = 1 \\ \frac{6}{19} & \text{if } x_2 = 2 \end{cases} \quad (\text{simplifying})$$

11. Consider the joint PDF:

$$f_{Y_1 Y_2}(y_1, y_2) = 3y_1, \quad 0 \leq y_2 \leq y_1 \leq 1$$

(a) Find the marginal PDFs for Y_1 and Y_2 .

To find the marginal PDFs, we integrate over each of the variables:

$$f_{Y_1}(y_1) = \int_0^{y_1} 3y_1 \, dy_2 \quad (\text{integrating out } y_2)$$

$$= 3y_1 y_2 \Big|_0^{y_1} \quad (\text{integrating})$$

$$f_{Y_1}(y_1) = 3y_1^2, \quad y_1 \in (0, 1) \quad (\text{the marginal PDF of } y_1)$$

$$f_{Y_2}(y_2) = \int_{y_2}^1 3y_1 \, dy_1 \quad (\text{integrating out } y_1)$$

$$= \frac{3}{2} y_1^2 \Big|_{y_2}^1 \quad (\text{integrating})$$

$$f_{Y_2}(y_2) = \frac{3}{2} - \frac{3}{2} y_2^2, \quad y_2 \in (0, 1) \quad (\text{the marginal PDF of } y_2)$$

- (b)
- Find the conditional PDF of Y_2 given $Y_1 = y_1$.**

Again, we can find the conditional PDF by taking the ratio of the joint and marginal PDFs:

$$f(y_2|Y_1 = y_1) = \frac{3y_1}{3y_1^2} \quad (\text{the ratio } f(y_1, y_2)/f(y_1))$$

$$f(y_2|Y_1 = y_1) = \frac{1}{y_1}, \quad y_2 \in (0, y_1) \quad (\text{the conditional PDF})$$

- (c)
- What is the probability that $Y_2 \geq 1/2$ given that $Y_1 = 3/4$?**

To find the conditional probability, we can integrate over the relevant range using the conditional PDF:

$$P(Y_2 \geq 1/2|Y_1 = 3/4) = \int_{1/2}^{3/4} \frac{1}{3/4} dy_2 \quad (\text{integrating over the cond. PDF})$$

$$= \frac{4}{3} y_2 \Big|_{1/2}^{3/4} \quad (\text{integrating})$$

$$P(Y_2 \geq 1/2|Y_1 = 3/4) = \frac{1}{3} \quad (\text{evaluating})$$

12. **For each of the following joint PDFs, determine whether or not the random variables x and y are independent.**

- (a)
- $f_{XY}(x, y) = 1$
- ,
- where $x \in [0, 1]$ and $y \in [0, 1]$**

Again, we know that these random variables are independent by inspection; this is a uniformly distribution over a square. More formally, we can factor the PDF into the product of two functions:

$$f_{XY}(x, y) = (\mathbb{1}\{0 \leq x \leq 1\}) (\mathbb{1}\{0 \leq y \leq 1\})$$

Each indicator is a function solely of x or solely of y ; thus the random variables are independent.

- (b)
- $f_{XY}(x, y) = e^{-(x+y)}$
- ,
- where $x > 0$ and $y > 0$**

Again, we can factor this joint PDF into the product of two functions:

$$f_{XY}(x, y) = (e^{-x}) (e^{-y})$$

Since each is a function solely of x or solely of y , the random variables are independent.

- (c)
- $f_{XY}(x, y) = 6(1 - y)$
- ,
- where $0 \leq x \leq y \leq 1$**

We know immediately that these random variables are NOT independent, because we have dependence in the support. There are other instances, however, where a more rigorous explanation would be necessary. To show (more formally) that X and Y are not independent, we can find the marginal distributions. The marginal distribution of X :

$$f_X(x) = \int_x^1 6(1 - y) dy \quad (\text{integrating out the } y\text{s})$$

$$= 6y - 3y^2 \Big|_x^1 \quad (\text{integrating})$$

$$f_X(x) = 3x^2 - 6x + 3, \quad x \in [0, 1] \quad (\text{the marginal PDF of } x)$$

Finding the marginal distribution of Y :

$$f_Y(y) = \int_0^y 6(1-y) dx \quad (\text{integrating out the } xs)$$

$$= 6(1-y)x \Big|_0^y \quad (\text{integrating})$$

$$f_Y(y) = 6y - 6y^2, \quad y \in [0, 1] \quad (\text{the marginal PDF of } y)$$

If X and Y are independent, then the joint PDF needs to be the product of the marginals. In this case, however:

$$6(1-y) \neq (3x^2 - 6x + 3)(6y - 6y^2)$$

Thus, X and Y are not independent.

13. **Show that** $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - \mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \quad (\text{expanding})$$

$$= \mathbb{E}[XY] - \mathbb{E}[\mathbb{E}[X]Y] - \mathbb{E}[X\mathbb{E}[Y]] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]] \quad (\text{by linearity of } \mathbb{E})$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \quad (\text{pulling out constants})$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (\text{simplifying})$$

14. **Let** X_1, \dots, X_n **be a random sample from a distribution with mean** μ **and** $\sigma^2 < \infty$. **Show that:**

(a) $\mathbb{E}[\bar{X}] = \mu$

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \quad (\text{by def. of } \bar{X})$$

$$= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \quad (\text{distributing the expectation})$$

$$= \frac{1}{n} \sum_{i=1}^n \mu \quad (\text{plugging in the mean})$$

$$\mathbb{E}[\bar{X}] = \mu \quad (\text{simplifying})$$

(b) $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \quad (\text{by def. of } \bar{X})$$

$$= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \quad (\text{pulling out the constant})$$

$$= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \quad (\text{by independence})$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \quad (\text{plugging in the variance})$$

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad (\text{simplifying})$$

$$\begin{aligned}
\text{(c)} \quad \sum_{i=1}^n (X_i - \bar{X})^2 &= \left(\sum_{i=1}^n X_i^2 \right) - n\bar{X}^2 \\
&= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) && \text{(expanding)} \\
&= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 && \text{(distributing the sum)}
\end{aligned}$$

Note that $\sum X_i = n\bar{X}$. This lets us simplify the second term in the expression:

$$\begin{aligned}
&= \sum_{i=1}^n X_i^2 - 2\bar{X}[n\bar{X}] + n\bar{X}^2 && \text{(simplifying)} \\
\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n X_i^2 - n\bar{X}^2
\end{aligned}$$

$$\text{(d)} \quad \mathbb{E}[S^2] = \sigma^2$$

$$\begin{aligned}
\mathbb{E}[S^2] &= \mathbb{E} \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right] && \text{(by def. of } S^2) \\
&= \mathbb{E} \left[\frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \right] && \text{(expanding)} \\
&= \mathbb{E} \left[\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \right) \right] && \text{(distributing the sum)}
\end{aligned}$$

Note that $\sum X_i = n\bar{X}$. This lets us simplify the second term in the expectation:

$$\begin{aligned}
&= \mathbb{E} \left[\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \right) \right] && \text{(simplifying)} \\
&= \mathbb{E} \left[\frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right) \right] && \text{(rearranging)} \\
&= \left(\frac{n}{n-1} \right) \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}^2] \right) && \text{(distributing the expectation)}
\end{aligned}$$

Recall that $\mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}[X]^2$. Thus, we can plug in for the expectations of $\mathbb{E}[X_i^2]$ and $\mathbb{E}[\bar{X}^2]$, using the results in (a) and (b):

$$\begin{aligned}
&= \left(\frac{n}{n-1} \right) \left(\frac{1}{n} \sum_{i=1}^n [\sigma^2 + \mu^2] - \left[\frac{\sigma^2}{n} + \mu^2 \right] \right) && \text{(plugging in)} \\
&= \left(\frac{n}{n-1} \right) \left(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} + \mu^2 \right) && \text{(taking the sum)} \\
&= \left(\frac{n}{n-1} \right) \left(\frac{(n-1)\sigma^2}{n} \right) && \text{(simplifying)} \\
\mathbb{E}[S^2] &= \sigma^2
\end{aligned}$$

15. Let Y_1, \dots, Y_n be a random sample from a distribution with PDF

$$f(y_i|\theta) = \begin{cases} (\theta + 1)y^{-(\theta+2)} & \text{if } y > 1 \\ 0 & \text{else} \end{cases}$$

where $\theta \in (0, \infty)$.

(a) Find the method of moments estimator for θ .

Finding the population moment:

$$\mathbb{E}[X] = \int_1^\infty x(\theta + 1)x^{-\theta-2} dx \quad (\text{the expected value})$$

$$= \int_1^\infty (\theta + 1)x^{-\theta-1} dx \quad (\text{simplifying})$$

$$= \left[-\frac{\theta + 1}{\theta} x^{-\theta} \right]_1^\infty \quad (\text{taking the integral})$$

$$\mathbb{E}[X] = \frac{\theta + 1}{\theta} \quad (\text{evaluating})$$

$$M_1 = 1 + \frac{1}{\theta} \quad (\text{the first pop. moment})$$

We can use our usual sample moment and match it with the population moment to find our estimator:

$$m_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{the first sample moment})$$

$$\frac{1}{n} \sum_{i=1}^n X_i = 1 + \frac{1}{\hat{\theta}} \quad (\text{matching moments})$$

$$\hat{\theta} = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i - 1} \quad (\text{solving for } \hat{\theta})$$

$$\hat{\theta}_{mm} = \frac{1}{\bar{X} - 1} \quad (\text{simplifying})$$

(b) Find the maximum likelihood estimator for θ .

Again, because we have an iid sample, we can simply multiply the PDFs of each draw together to find the likelihood functions:

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{i=1}^n (\theta + 1)x_i^{-\theta-2} \quad (\text{the likelihood})$$

$$= (\theta + 1)^n \prod_{i=1}^n x_i^{-\theta-2} \quad (\text{multiplying})$$

$$\ln(\mathcal{L}(\cdot)) = n \ln(\theta + 1) - (\theta + 2) \sum_{i=1}^n \ln(x_i) \quad (\text{taking the log})$$

$$\frac{\partial \ln(\mathcal{L}(\cdot))}{\partial \theta} = \frac{n}{\theta + 1} - \sum_{i=1}^n \ln(x_i) = 0 \quad (\text{the FOC})$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \ln(x_i)} - 1 \quad (\text{solving for } \hat{\theta})$$

$$\hat{\theta}_{mle} = \frac{n}{\sum_{i=1}^n \ln(X_i)} - 1 \quad (\text{simplifying})$$

16. For each of the following PDFs, let X_1, \dots, X_n be a random sample. Find a complete sufficient statistic for the unknown parameter(s).

(a) $f(x|\theta) = \frac{\theta}{(1+x)^{1+\theta}}$, where $0 < x < \infty$ and $\theta > 0$

$$f(x|\theta) = \theta(1+x)^{-1}(1+x)^{-\theta} \quad (\text{rearranging the PDF})$$

$$= (1+x)^{-1}\theta \exp\{-\theta \ln(1+x)\} \quad (\text{rearranging further})$$

From our exponential family definition, let $h(x) = (1+x)^{-1}$, let $c(\theta) = \theta$, let $w(\theta) = -\theta$, and let $t(x) = \ln(1+x)$. Thus, this distribution belongs to the exponential family. This gives us the statistic $t(x)$:

$$t(x) = \ln(1+x) \quad (\text{the statistic in the exponent})$$

By our exponential family theorem, the statistic $T(\mathbf{x}) = \sum t(x_i)$ is complete:

$$T(\mathbf{x}) = \sum_{i=1}^n \ln(1+x_i) \quad (\text{a complete statistic})$$

Further, we can show that $T(\mathbf{x})$ is sufficient:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n (1+x_i)^{-1}\theta \exp\{-\theta \ln(1+x_i)\} \quad (\text{the joint PDF})$$

$$= \left(\prod_{i=1}^n \frac{1}{1+x_i} \right) \theta^n \exp\left\{ \theta \sum_{i=1}^n \ln(1+x_i) \right\} \quad (\text{multiplying})$$

Now, letting $g(T(\mathbf{x})|\theta) = \theta^n \exp\{\theta \sum \ln(1+x_i)\}$ from our factorization theorem, we know that $T(\mathbf{x})$ is also sufficient:

$$T(\mathbf{x}) = \sum_{i=1}^n \ln(1+x_i) \quad (\text{by the factorization theorem})$$

(b) $f(x|\beta) = \frac{\ln[\beta]\beta^x}{\beta-1}$, where $0 < x < 1$ and $\beta > 1$

$$f(x|\beta) = \frac{\ln(\beta)}{\beta-1} \exp\{\ln(\beta)x\} \quad (\text{rearranging the PDF})$$

Again, this distribution belongs the exponential family. Let $h(x) = 1$, let $c(\beta) = \ln(\beta)(\beta-1)^{-1}$, let $w(\beta) = \ln(\beta)$, and let $t(x) = x$. By the same logic above (we could employ the factorization theorem, which will always produce the same statistic we found using our completeness theorem), our complete sufficient statistic is:

$$T(\mathbf{x}) = \sum_{i=1}^n x_i$$

(c) $f(x|\theta) = \exp\{-(x-\theta)\} \exp\{-e^{-(x-\theta)}\}$, where $-\infty < x < \infty$ and $-\infty < \theta < \infty$

$$f(x|\theta) = \exp\{-x\} \exp\{\theta\} \exp\{-e^\theta e^{-x}\} \quad (\text{rearranging the PDF})$$

Yet again, this distribution belongs to the exponential family, with $h(x) = e^{-x}$ and $c(\theta) = e^{-\theta}$. Further, $w(\theta) = -e^\theta$, and $t(x) = e^{-x}$. Thus, our complete sufficient statistic is

$$T(\mathbf{x}) = \sum_{i=1}^n \exp \{ -x_i \}$$

(d) $f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$ **where** $0 < x < 1$ **and** $\alpha, \beta > 0$

There are two parameters for this distribution, so we will be after a bivariate statistic:

$$\begin{aligned} f(x|\alpha, \beta) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} && \text{(rearranging the PDF)} \\ &= \left(\frac{1}{x(1-x)} \right) \left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right) \exp \{ \alpha \ln(x) + \beta \ln(1-x) \} && \text{(rearranging further)} \end{aligned}$$

Now, we let $h(x) = [x(1-x)]^{-1}$; let $c(\alpha, \beta) = \Gamma(\alpha + \beta)\Gamma(\alpha)^{-1}\Gamma(\beta)^{-1}$; let $w_1(\alpha, \beta) = \alpha$ and $w_2(\alpha, \beta) = \beta$; and let $t_1(x) = \ln(x)$ and $t_2(x) = \ln(1-x)$. Thus, we have a bivariate, complete, sufficient statistic:

$$T(\mathbf{x}) = \begin{pmatrix} \sum_{i=1}^n \ln(x_i) \\ \sum_{i=1}^n \ln(1-x_i) \end{pmatrix}$$

Note that these two statistics are jointly sufficient for α and β .