Models of course, are never true but fortunately it is only necessary that they be useful. –George Box (1979)

Exercise is the beste instrument in learnyng. –Robert Recorde, *The Wheststone of Witte* (1557)

Le juge: Accusé, vous tâcherez d'être bref. *L'accusé*: Je tâcherai d'être clair. [Judge: Defendant, you should be brief. Defendant: I must be clear.] –G. Courteline

Solutions to Problems 91-100 More Re-enforcements

91. Let X_1 and X_2 be independent standard normal random variables. Let $U \sim U(0,1)$ be independent of X_1 and X_2. Define $Z = UX_1 + (1-U)X_2$. (a) Find the conditional distribution of $Z|(U=u)$. (b) Find $\mathbb{E}(Z)$ and $\mathbb{V}(Z)$.

Solution. (a) By the well known result of 'independent linear combination preserves normality' we know that

$$Z|(U=u) \sim N(u\mu_1 + (1-u)\mu_2, u^2\sigma_1^2 + (1-u)^2\sigma_2^2) = N(0, u^2 + (1-u)^2)$$

as $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$.

(b) $\mathbb{E}(Z) = \mathbb{E}[\mathbb{E}[Z|U]] = \mathbb{E}[0] = 0$,

$$\mathbb{V}(Z) = \mathbb{E}\big[\mathbb{V}[Z|U]\big] + \mathbb{V}\big[\mathbb{E}[Z|U]\big] = \mathbb{E}[U^2] + \mathbb{E}[(1-U)^2] + \mathbb{V}[0] = \frac{1}{3} + \frac{1}{3} + 0 = \frac{2}{3}$$

92. For a random variable $X \sim N(0,1)$, define $Y = X^2$. (a) Find the distribution of Y. (b) Find the correlation coefficient of X and Y. Are X and Y independent? (c) Find the conditional distribution of $Y|(X=x)$ and $X|(Y=y)$. (d) Find the joint cumulative distribution function of X and Y. (e) Can the identities "$f(x,y) = f(x)f(y|x) = f(y)f(x|y)$" be used to derive the joint density function of X and Y? Comment on the existence of the joint density function of X and Y in the two dimensional space.

Solution. (a) with the symmetry of $N(0,1)$ in mind, we write $f_Y(y)dy = f_X(x)dx\mathbb{I}(X>0) + f_X(x)dx\mathbb{I}(X<0)$=>

$$f_Y(y) = 2f_X(x) \cdot \frac{1}{2x}\mathbb{I}(x>0) = \frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}\mathbb{I}(x>0) = \frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}\mathbb{I}(Y>0)$$

(b) $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$, $\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$, $\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\mathbb{V}(X)\mathbb{V}(Y)} = 0$. But clearly Y is <u>functionally determined</u> by X via the relation $Y = X^2$ and therefore they cannot be independent.

(c) The density of $Y|(X=x)$ is concentrated at $y = x^2$, that is,

$$f_{Y|X}(y|x) = \mathbb{I}(y = x^2)$$

The density of $X|(Y=y)$ is concetrated and equally divided between $x=\pm\sqrt{y}$, that is,

$$f_{X|Y}(x|y)=\frac{1}{2}\mathbb{I}(x=\sqrt{y})+\frac{1}{2}\mathbb{I}(x=-\sqrt{y})$$

(d) The trick is to think of the graph and fix Y at the positive y but let x vary. There turns out to exist only 3 different cases.

$$F_{(X,Y)}(x,y):=\mathbb{P}(X\leq x,Y\leq y)=\begin{cases}0 & x\leq-\sqrt{y}\\ \Phi(x)-\Phi(-\sqrt{y}) & -\sqrt{y}<x\leq\sqrt{y}\\ \Phi(\sqrt{y})-\Phi(-\sqrt{y}) & x>\sqrt{y}\end{cases}$$

Simplyfing to

$$F_{(X,Y)}(x,y)=\left[\Phi\left(\min(x,\sqrt{y})\right)-\Phi(-\sqrt{y})\right]\mathbb{I}(-\sqrt{y}<x)\mathbb{I}(y>0)$$

(e)

$$f_{Y|X}(y|x)f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathbb{I}(y=x^2)$$

$$f_{X|Y}(x|y)f_Y(x)=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}\left[\frac{1}{2}\mathbb{I}(x=\sqrt{y})+\frac{1}{2}\mathbb{I}(x=-\sqrt{y})\right]$$

But the two expressions do not equal to each other, e.g., at $(x,y)=(1,1)$. This means the formula breaks down in this "singular" case when the joint density surface has "vertical holes".

If we make strict that the joint density is defined only as the derivative of the cumulative distribution function, then, in event of discontinuity, such derivative does not exist. (Recall that a function is discontinuous only if its graph vertially broken. Discountinous domain, or any horizontal break, does not result in a discontinous function defined on that domain.)

93. Suppose that the conditional density of the random vector (X,Y) given the random variable Z is $f(x,y|z)=[z+(1-z)(x+y)]I_{(0,1)}(x)I_{(0,1)}(y)$ for $0\leq z\leq 2$, and the density of the random variable Z is $f(z)=\frac{1}{2}I_{[0,2]}(z)$ where $I_A(x)$ denotes the indicator function, i.e., $I_A(x)=1$ iff $x\in A$ and $I_A(x)=0$ iff $x\notin A$. (a) Find the expectation $\mathbb{E}(X+Y)$. (b) Determine whether X and Y are independent or not. (c) Determine whether X and Z are independent or not. (d) Find the joint density of X and $X+Y$. (e) Find the distribution function of $\max(X,Y)\,|(Z=z)$.

Solution. (a) $\mathbb{E}(X+Y)=\mathbb{E}[\mathbb{E}(X+Y|Z)]=\int_0^2 \frac{1}{2}\int_0^1\int_0^1(x+y)[z+(1-z)(x+y)]dxdy\,dz=\frac{1}{2}\int_0^2 zdz\int_0^1\int_0^1(x+y)dxdy+\frac{1}{2}\int_0^2(1-z)dz\int_0^1\int_0^1(x+y)^2dxdy=\frac{1}{2}\cdot 2\cdot\left(\frac{1}{2}+\frac{1}{2}\right)+0=1$

(b) $f_{(X,Y)}(x,y)=\mathbb{E}\left[f_{(X,Y|Z)}(x,y|Z)\right]=\int_0^2\frac{1}{2}[z+(1-z)(x+y)]I_{(0,1)}(x)I_{(0,1)}(y)\,dz=\frac{1}{2}I_{(0,1)}(x)I_{(0,1)}(y)\left[\int_0^2 zdz+(x+y)\int_0^2 dz-(x+y)\int_0^2 zdz\right]$

$$=\frac{1}{2}I_{(0,1)}(x)I_{(0,1)}(y)[2+2(x+y)-2(x+y)]=I_{(0,1)}(x)I_{(0,1)}(y)$$

Clearly, the factorization of the joint density of (X,Y) into one part independent of y and the other part independent of x means that the two r.v.s are independent of each other and that $f_X(x)=I_{(0,1)}(x)$ and $f_Y(y)I_{(0,1)}(y)$ are the marginal densities.

(c) $f_{(X,Z)}(x,z)=\int_0^1 f(x,y|z)f(z)dy=\int_0^1\frac{1}{2}[z+(1-z)(x+y)]I_{(0,1)}(x)I_{(0,2)}(z)dy=[\frac{1}{4}(1+z)+\frac{1}{2}(1-z)x]I_{(0,1)}(x)I_{(0,2)}(z)$ which does not factorize over x and z, implying that X and Z are not independent.

(d) Differential linearity gives $d\begin{bmatrix}x\\s\end{bmatrix}=\begin{bmatrix}1&0\\1&1\end{bmatrix}d\begin{bmatrix}x\\y\end{bmatrix}\Rightarrow\frac{dxds}{dxdy}=|\det\begin{bmatrix}1&0\\1&1\end{bmatrix}|=1$. Measurability gives $f_{(X,S)}(x,s)dxds=f_{(X,Y)}(x,y)dxdy$

$$\therefore\ f_{(X,S)}(x,s)=f_{(X,Y)}(x,y)=I_{(0,1)}(x)I_{(0,1)}(y)=I_{(0,1)}(x)I_{(0,1)}(s-x)$$

(e) $\mathbb{P}(\max(X,Y)\le t|Z=z)=\mathbb{P}(X\le t,Y\le t|Z=z)=\int_{-\infty}^t\int_{-\infty}^t f(x,y|z)dxdy$

$$=\begin{cases}0 & t<0\\ zt^2+(1-z)t^3 & 0\le t\le 1\\ 1 & t>1\end{cases}$$

$$=\min\left([zt^2+(1-z)t^3]I_{[0,1]}(t),1\right)$$

which gives the cumulative distribution function of the r.v. $\max(X,Y)\,|Z=z$. The final one-liner simplificatio is due to the fact that $zt^2+(1-z)t^3\ge t^2$ when $t\ge 1$.

94. Let $a_1,\dots,a_n$ be positive constants and $b_1,\dots,b_{n-1}$ be nonnegative constants. Suppose that random variables $Y_1,\dots,Y_{n-1}$ are mutually independent and $Y_j\sim\text{Beta}\left(d_j,a_{j+1}\right)$, where $d_j=\sum_{k=1}^j(a_k+b_k)$, $j=1,\dots,n-1$. Define

$$\begin{cases} X_i = (1 - Y_{i-1}) \prod_{j=i}^{n-1} Y_j, & i = 1, \dots, n-1 \\ X_n = 1 - Y_{n-1} \end{cases}$$

where $Y_0 = 0$. Prove that the joint density of $X_1, \dots, X_{n-1}$ is given by

$$\frac{\prod_{i=1}^{n} x_i^{a_i - 1} \prod_{j=1}^{n-1} \left(\sum_{k=1}^{j} x_k\right)^{b_j}}{\prod_{j=1}^{n-1} \mathrm{B}\left(d_j, a_{j+1}\right)}$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function. [Hint: $X_1 + \cdots + X_i = Y_i Y_{i+1} \cdots Y_{n-1}, i = 1, \dots, n-1$.]

Solution. We start with explicitizing the $X = X(Y)$ expressions:

$$\begin{cases} X_1 & = & Y_1 Y_2 Y_3 \cdots Y_{n-1} \\ X_2 & = & (1 - Y_1) Y_2 Y_3 \cdots Y_{n-1} \\ X_3 & = & (1 - Y_2) Y_3 \cdots Y_{n-1} \\ & \vdots & \\ X_{n-1} & = & (1 - Y_{n-2}) Y_{n-1} \\ X_n & = & (1 - Y_{n-1}) \end{cases}$$

The joint density of Y is already known, we need the Jacobian $\frac{dy_1 \cdots dy_{n-1}}{dx_1 \cdots dx_{n-1}}$ so that we can use the identity given by measurability:

$$f_X(x) \cdot \prod dx_i = f_Y(y) \cdot \prod dy_i$$

to find $f_X(x)$, where unsubscripted letters represent vectors.

From the Hint, we see that

$$\frac{X_i}{X_1 + \cdots + X_i} = 1 - Y_{i-1} \Rightarrow Y_{i-1} = \frac{X_1 + \cdots + X_{i-1}}{X_1 + \cdots + X_i} = \frac{S_{i-1}}{S_{i-1} + X_i}$$

where

$$S_{i-1} = X_1 + \cdots + X_{i-1} = Y_{i-1} Y_i \cdots Y_{n-1}.$$

written explicitely as

$$\begin{array}{rcr} S_1 & = & Y_1 Y_2 Y_3 \cdots Y_{n-1} \\ S_2 & = & Y_2 Y_3 \cdots Y_{n-1} \\ S_3 & = & Y_3 \cdots Y_{n-1} \\ & \vdots & \\ S_{n-2} & = & Y_{n-2} Y_{n-1} \\ S_{n-1} & = & Y_{n-1} \end{array}$$

We can use S as a bridge: $\frac{\prod dy_i}{\prod dx_i} = \frac{\prod dy_i}{\prod ds_i} \cdot \frac{\prod ds_i}{\prod dx_i} = \frac{\prod ds_i}{\prod dx_i} \Big/ \frac{\prod ds_i}{\prod dy_i}$.

Observe that $\frac{\partial s_i}{\partial x_j} = \mathbb{I}(i \leq j)$, therefore $\frac{\prod ds_i}{\prod dx_i}$ is an upper triangular matrix with all non-zero elements equal to one. Thus the jacobian determinant is

$$\left|\det \frac{\prod ds_i}{\prod dx_i}\right| = 1$$

Observe that $\frac{\partial s_i}{\partial y_j} = c\mathbb{I}(i \geq j)$, therefore $\frac{\prod ds_i}{\prod dy_i}$ is a lower triangular matrix and its diagonal elements are

$$\begin{array}{rcr} \frac{\partial s_1}{\partial y_1} & = & y_2 y_3 \cdots y_{n-1} \\ \frac{\partial s_2}{\partial y_2} & = & y_3 \cdots y_{n-1} \\ & \vdots & \\ \frac{\partial s_{n-1}}{\partial y_{n-1}} & = & 1 \end{array}$$

Thus the jacobian determinant is

$$\left|\det \frac{\prod ds_i}{\prod dy}\right| = \prod_{i=2}^{n-1} y_j^{j-1} = \prod_{i=1}^{n-1} y_j^{j-1}.$$

Thus,

$$\left|\det \frac{\prod dy_i}{\prod dx_i}\right| = \frac{1}{\prod_{i=1}^{n-1} y_j^{j-1}}.$$

Next we explicitize $f_Y(y)$, the joint pdf of $(Y_1, Y_2, \ldots, Y_{n-1})$:

$$
\begin{aligned}
Y_1 &\sim \frac{1}{B(d_1,a_2)} y_1^{d_1-1}(1-y_1)^{a_2-1}\\
Y_2 &\sim \frac{1}{B(d_2,a_3)} y_2^{d_2-1}(1-y_2)^{a_3-1}\\
&\vdots\\
Y_{n-1} &\sim \frac{1}{B(d_{n-1},a_n)} y_{n-1}^{d_{n-1}-1}(1-y_{n-1})^{a_n-1}
\end{aligned}
$$

$$
\therefore\ f_Y(y) = \frac{\prod_{j=1}^{n-1} y_j^{d_j-1} \prod_{j=1}^{n-1} \left(1-y_{j-1}\right)^{a_j-1}}{\prod_{j=1}^{n-1} B\left(d_j, a_{j+1}\right)}
$$

Hence

$$
f_X(x) = \frac{1}{\prod_{j=1}^{n-1} y_j^{j-1}} \cdot \frac{\prod_{j=1}^{n-1} y_j^{d_j-1} \prod_{j=1}^{n-1} \left(1-y_{j-1}\right)^{a_j-1}}{\prod_{j=1}^{n-1} B\left(d_j, a_{j+1}\right)} = \frac{\prod_{j=1}^{n-1} y_j^{d_j-j} \prod_{i=1}^{n-1} \frac{x_i^{a_i-1}}{S_i^{a_i-1}}}{\prod_{j=1}^{n-1} B\left(d_j, a_{j+1}\right)} \overset{!!}{=} \frac{\prod_{i=1}^{n-1} x_i^{a_i-1}}{\prod_{j=1}^{n-1} B\left(d_j, a_{j+1}\right)} \cdot \frac{\prod_{j=1}^{n-1} y_j^{d_j-j}}{\prod_{i=1}^{n-1} y_i^{\sum_{k=1}^{i}(a_k-1)}}
$$

$$
= \frac{\prod_{i=1}^{n-1} x_i^{a_i-1} \prod_{j=1}^{n-1} y_j^{\sum_{k=1}^{j} b_k}}{\prod_{j=1}^{n-1} B\left(d_j, a_{j+1}\right)} \overset{!!}{=} \frac{\prod_{i=1}^{n-1} x_i^{a_i-1} \prod_{j=1}^{n-1} S_j^{b_j}}{\prod_{j=1}^{n-1} B\left(d_j, a_{j+1}\right)}
$$

where we have use the following identity twice:

$$
\prod_{i=1}^{n-1} S_i^{r_i} = \prod_{i=1}^{n-1} Y_i^{\sum_{k=1}^{i} r_k}.
$$

95. Let X_1 and X_2 be two independent standard normal random variables. Define $Y_1 = X_1 + X_2$ and $Y_2 = X_1/X_2$. (a) Find the joint density of Y_1 and Y_2. (b) Find the marginal density of Y_2.

Solution.

(a)

$$d\begin{bmatrix}y_1\\y_2\end{bmatrix}=\begin{bmatrix}1 & 1\\ \dfrac{1}{x_2} & -\dfrac{x_1}{x_2^2}\end{bmatrix}d\begin{bmatrix}x_1\\x_2\end{bmatrix}\Rightarrow dy_1dy_2=\left|\det\begin{bmatrix}1 & 1\\ \dfrac{1}{x_2} & -\dfrac{x_1}{x_2^2}\end{bmatrix}\right|dx_1dx_2=\frac{x_1+x_2}{x_2^2}dx_1dx_2$$

$$f_{(Y_1,Y_2)}(y_1,y_2)dy_1dy_2=f_{(X_1,X_2)}(x_1,x_2)dx_1dx_2\Rightarrow f_{(Y_1,Y_2)}(y_1,y_2)=f_{(X_1,X_2)}(x_1,x_2)\frac{x_2^2}{x_1+x_2}$$

$$x_1=y_2x_2\to y_1=y_2x_2+x_2\to x_2=\frac{y_1}{1+y_2}\to x_1=\frac{y_1y_2}{1+y_2}$$

$$\Rightarrow f_{(Y_1,Y_2)}(y_1,y_2)=\frac{1}{2\pi}e^{-\frac{x_1^2+x_2^2}{2}}\frac{x_2^2}{x_1+x_2}=\frac{1}{2\pi}e^{-\frac{y_1^2(1+y_2^2)}{2(1+y_2)^2}}\frac{y_1}{(1+y_2)^2}$$

(b)

$$f_{Y_2}(y_2)=\int_{-\infty}^{+\infty}f_{(Y_1,Y_2)}(y_1,y_2)dy_1=\frac{1}{2\pi(1+y_2)^2}\int_{-\infty}^{+\infty}e^{-\frac{1+y_2^2}{2(1+y_2)^2}y_1^2}y_1dy_1=\frac{1}{2\pi(1+y_2)^2}\int_{0}^{+\infty}e^{-\frac{1+y_2^2}{2(1+y_2)^2}t}dt=\frac{1}{2\pi(1+y_2)^2}\frac{2(1+y_2)^2}{1+y_2^2}=\frac{1}{\pi(1+y_2^2)}$$

which is a Cauchy density.

96. Let $X_1,\dots,X_n$ be a random sample from the uniform distribution on the unit interval $(0,1)$. The cumulative distribution function (cdf) of $X\sim U(0,1)$ is given by

$$\begin{cases}0, & \text{iff } x\le 0\\ x, & \text{iff } 0<x<1\\ 1, & \text{iff } x\ge 1\end{cases}$$

(a) Find the cdf and the density of $X_{(1)}=\min\{X_1,\dots,X_n\}$. (b) Find the cdf and the density of $X_{(n)}=\max\{X_1,\dots,X_n\}$. (c) Find the expression of $\mathbb{E}\left[X_{(n)}-X_{(1)}\right]$. (d) When $n=2$, find the value of $\mathbb{V}\left[X_{(n)}-X_{(1)}\right]$. [Hint: If $Y\sim\text{Beta}(a,b)$ then $\mathbb{E}(Y)=\frac{a}{a+b}$, $\mathbb{E}(Y^2)=\frac{a(a+1)}{(a+b)(a+b+1)}$, $\mathbb{V}(Y)=\frac{ab}{(a+b)^2(a+b+1)}$.]

Solution. (a) $\mathbb{P}\left(X_{(1)}>t\right)=\mathbb{P}(X_1>t,X_2>t,\cdots,X_n>t)=\mathbb{P}(X_1>t)\mathbb{P}(X_2>t)\cdots\mathbb{P}(X_n>t)=\begin{cases}1 & t\le 0\\ (1-t)^n & 0<t<1\\ 0 & t\ge 1\end{cases}$

Hence $\mathbb{P}(X_{(1)} \le t) = 1 - \mathbb{P}(X_{(1)} > t) = \begin{cases} 0 & t \le 0 \\ 1-(1-t)^n & 0 < t < 1 \\ 1 & t \ge 1 \end{cases}$

$f_{X_{(1)}}(t) = \frac{d}{dt}\mathbb{P}\big(X_{(1)} \le t\big) = n(1-t)^{n-1}\mathbb{I}(0<t<1)$

(b) $\mathbb{P}\big(X_{(n)} \le t\big) = \mathbb{P}(X_1 \le t, X_2 \le t, \cdots, X_n \le t) = \mathbb{P}(X_1 \le t)\mathbb{P}(X_2 \le t)\cdots\mathbb{P}(X_n \le t) = \begin{cases} 0 & t \le 0 \\ t^n & 0<t<1 \\ 1 & t \ge 1 \end{cases}$

$f_{X_{(n)}}(t) = \frac{d}{dt}\mathbb{P}\big(X_{(n)} \le t\big) = nt^{n-1}\mathbb{I}(0<t<1)$

(c) $\mathbb{E}\big[X_{(n)} - X_{(1)}\big] = \mathbb{E}\big[X_{(n)}\big] - \mathbb{E}\big[X_{(1)}\big] = \frac{n-1}{n+1}$ because

$\mathbb{E}[X_{(n)}] = \int_0^1 tnt^{n-1}dt = \frac{n}{n+1}$ and $\mathbb{E}\big[X_{(1)}\big] = \int_0^1 tn(1-t)^{n-1}dt = \int_0^1 (1-s)ns^{n-1}ds = \int_0^1 ns^{n-1}ds - \int_0^1 ns^n ds = 1 - \frac{n}{n+1} = \frac{1}{n+1}$

(d) $\mathbb{V}\big[X_{(2)} - X_{(1)}\big] = \mathbb{E}\big[X_{(2)}^2 + X_{(1)}^2 - 2X_{(1)}X_{(2)}\big] - \left(\frac{2-1}{2+1}\right)^2$

$\mathbb{E}[X_{(2)}^2] = \int_0^1 2t^3 dt = \frac{1}{2}$ and $\mathbb{E}[X_{(1)}^2] = \int_0^1 2t^2(1-t)dt = \frac{1}{6}$ and $\mathbb{E}[X_{(1)}X_{(2)}] = \mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{1}{4}$

$$\therefore \mathbb{V}\big[X_{(2)} - X_{(1)}\big] = \frac{1}{2} + \frac{1}{6} - \frac{2}{4} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

(Remark: The beta hint isn’t used.)

97. Let $X_1, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$. (a) Let $\sigma^2 = 1$. Find the most powerful test of size α for testing the null hypothesis $H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu = \mu_1 (< \mu_0)$. (b) Let μ be unknown. Use the likelihood ratio (LR) method to find the LR test of size α for testing $H_0: \sigma^2 = \sigma_0^2$ agianst the alternative $H_1: \sigma^2 \ne \sigma_0^2$.

Solution. (a) The hypotheses is of simple-vs-simple type ➔ The Neyman-Pearson Lemma is in effect.

$L_0 = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (X_i - \mu_0)^2}$ and $L_1 = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (X_i - \mu_1)^2}$

$k_1 \geq \frac{L_0}{L_1} = e^{-\frac{1}{2}\sum_{i=1}^{n}[(X_i-\mu_0)^2-(X_i-\mu_1)^2]}$

$k_2 \leq \sum_{i=1}^{n}[(X_i-\mu_0)^2-(X_i-\mu_1)^2] = \sum_{i=1}^{n}[(2X_i-\mu_0-\mu_1)(\mu_1-\mu_0)]$

$k_3 \geq \sum_{i=1}^{n}(2X_i-\mu_0-\mu_1)$

$k_4 \geq \sum_{i=1}^{n}X_i \overset{H_0}{\sim} N(\mu_0, n)$

Hence we set $k_4 = \mu_0 + n \cdot z_{1-\alpha}$ where $z_{1-\alpha}$ is the lower α quantile of the standard normal distribution to control the size of the test at α.

(b) Since μ is unknown, we have to replace it with its MLE. **Lemma**. The MLE for μ is $\bar{X}$ and the MLE for (μ, σ^2) is $\left(\bar{X}, \frac{(n-1)S^2}{n}\right)$. **Proof of Lemma**.
$L(\mu,\sigma^2;X) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2} \Rightarrow \ln L(\mu,\sigma^2;X) = -\frac{n}{2}(\ln 2\pi + \ln\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \Rightarrow \frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 \Rightarrow \hat{\mu} = \bar{X}$. One the other hand, $\frac{\partial \ln L}{\partial(\sigma^2)} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(X_i-\mu)^2 = 0 \Rightarrow \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$, in which we replace μ by $\hat{\mu} = \bar{X}$ to yield $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{(n-1)S^2}{n}$. **End of Proof of Lemma**.

Next we construct the (Generalized) Likelihood Ratio Test's Critical region:

$$L_0 = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$\max L = \left(\frac{1}{\sqrt{2\pi\frac{(n-1)S^2}{n}}}\right)^n e^{-\frac{n}{2(n-1)S^2}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$k_1 \geq \Lambda = \frac{L_0}{\max L} = \left(\frac{(n-1)S^2}{n\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\bar{X})^2\left(\frac{1}{\sigma_0^2}-\frac{n}{(n-1)S^2}\right)} = \left(\frac{(n-1)S^2}{n\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}(n-1)S^2\left(\frac{1}{\sigma_0^2}-\frac{n}{(n-1)S^2}\right)}$$

$$k_2 \geq \left(\frac{(n-1)S^2}{\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\frac{(n-1)S^2}{\sigma_0^2}}$$

$$T = \frac{(n-1)S^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2(n-1)$$

$$k_2 \geq T^{\frac{n}{2}} e^{-\frac{1}{2}T}$$

$$k_3 \geq n\ln T - T$$

It is easy to verify that $n\ln T - T$ is a concave function with a maximum at $T = n$. Knowing about the convex shape of $n\ln T - T$ allows us to see that the critical region on the range of T has two components:

$$k_3 \geq n\ln T - T \Leftrightarrow T \in (-\infty, c_1] \cup [c_2, +\infty)$$

Next we determine the boundaries $c_{1,2}$ by solving the following two equaltions for c_1 and c_2, the only two unknowns:

$$\begin{cases} \text{ChisqPDF}[c_1;\ \text{df} = n-1] = \text{ChisqPDF}[c_2;\ df = n-1] \\ \text{ChisqCDF}[c_1;\ df = n-1] + 1 - \text{ChisqCDF}[c_2;\ \text{df} = n-1] = \alpha \end{cases} \xrightarrow{\text{solve for}} (c_1, c_2)$$

The solution is usually done by numerical iteration.

Alternatively, we may simplify the solution by relaxing the condition to “equal-tailed”, that is, imposing that the two components of the critical region to have equal probabilities under null hypothesis. Of course, this is at the cost of the power of the test. The equal-tailed boundaries are just the lower and upper-$\frac{\alpha}{2}$ quantile of the $\chi^2(n-1)$ distribution:

$$c_1^* = \chi^2_{1-\frac{\alpha}{2}}(n-1)$$
$$c_2^* = \chi^2_{\frac{\alpha}{2}}(n-1)$$

where χ^2_α denotes the upper-α quantile of the distribution.

98. Consider a random sample $X_1, \ldots, X_n$ from the Poisson distribution $f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!}, x = 0,1,2, \ldots$, where $\theta > 0$. (a) Use the likelihood ratio test to find the general form of the critical region C for testing $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$, where $\theta_0 > 0$. (b) Given the critical region $C = \{(x_1, \ldots, x_n): |\bar{x} - \theta_0| \geq k\}$, when $n \to \infty$, use the central limit theorem to find k, where the size is α.

Solution. (a) Since the null parameter space is a singleton and the alternative parameter space is the full parameter space except the null value, the (Generalized) Likelihood Ratio Test Statsitic is

$$\Lambda = \frac{L(\theta_0)}{\max L\,(\theta)}$$

And it is well known that for an i.i.d. Poisson(θ) sample, mle(θ)=$\bar{x}$. Thus the critical region based on the (Generalized) Likelihood Ratio Test is

$$\mathbb{C} = \left\{ \mathbf{X}: \Lambda = \frac{L(\theta_0)}{L(\bar{X})} \leq k_1 \right\}$$

Proceeding to computation:

$$L(\theta_0) = e^{-n\theta_0} \frac{\theta_0^{n\bar{X}}}{\prod_{i=1}^{n} X_i!}$$
$$L(\bar{X}) = e^{-n\bar{X}} \frac{\bar{X}^{n\bar{X}}}{\prod_{i=1}^{n} X_i!}$$
$$\Lambda = e^{-n(\theta_0 - \bar{X})} \left(\frac{\theta_0}{\bar{X}}\right)^{n\bar{X}}$$
$$k_1 \geq e^{-n(\theta_0 - \bar{X})} \left(\frac{\theta_0}{\bar{X}}\right)^{n\bar{X}}$$
$$k_2 \geq e^{n\bar{X}} \bar{X}^{-n\bar{X}}$$
$$k_3 \geq \bar{X} - \bar{X} \ln \bar{X}$$

Note that the Poisson distribution is independently additive, that is, the sample sum

$$n\bar{X} \overset{H_0}{\sim} \text{Poisson}(n\theta_0).$$

It is easy to see that $\bar{X} - \bar{X} \ln \bar{X}$ is concave and maximized at $\bar{X} = 1$. The knowledge about the concave shape allows us to see that the critical region on the range of $\bar{X}$ has 2 components:

$$k_3 \geq \bar{X} - \bar{X}\ln\bar{X} \Leftrightarrow n\bar{X} \in [0, c_1] \cup [c_2, +\infty)$$

That is, we will be using the sample sum

$$T = n\bar{X}$$

as the test statistic. Next we determine the boundaries $c_{1,2}$ by solving the following two equations:

$$\begin{cases} \text{PoissonPDF}[c_1;\ n\theta_0] = \text{PoissonPDF}[c_2;\ n\theta_0] \\ \text{PoissonCDF}[c_1; n\theta_0] + 1 - \text{PoissonCDF}[c_2;\ n\theta_0] = \alpha \end{cases} \xRightarrow{\text{solve for}} (c_1, c_2)$$

The system is usually solved by numerical iteration.

In case that <u>no</u> $c_1 \geq 0$ can satisfy the system, then we simply discard the $[0, c_1]$ component of the critical region, leaving it to be the one-sided $[c_2, +\infty)$ where $c_2 = \text{Poisson}_\alpha(\theta_0)$ is the upper-α quantile of the distribution.

Alternatively, we may simplify the solution by requiring that the two components to carry the same probability $\frac{\alpha}{2}$. Hence the equal-tail boundaries are

$$\begin{cases} c_1^* = \text{Poisson}_{1-\frac{\alpha}{2}}(n\theta_0) \\ c_2^* = \text{Poisson}_{\frac{\alpha}{2}}(n\theta_0) \end{cases}$$

In case that no $c_1^* \geq 0$ can satisfy the first expression, then we simply discard the $[0, c_1^*]$ component of the critical region, leaving it to be the one-sided $[c_2^*, +\infty)$ where $c_2^* = \text{Poisson}_\alpha(\theta_0)$ is the upper-α quantile of the distribution.

(b) Since Poisson is independently additive, it has normal approximation.

$$n\bar{X} \to N(n\theta_0, n\theta_0)$$
$$\frac{n\bar{X} - n\theta_0}{\sqrt{n\theta_0}} \xrightarrow{n\to\infty} N(0,1)$$
$$|\bar{X} - \theta_0| \geq k \Leftrightarrow \left|\frac{n\bar{X} - n\theta_0}{\sqrt{n\theta_0}}\right| \geq k\sqrt{\frac{n}{\theta_0}}$$

$$|\bar{X} - \theta_0| \geq k \Leftrightarrow \left|\frac{n\bar{X} - n\theta_0}{\sqrt{n\theta_0}}\right| \geq k\sqrt{\frac{n}{\theta_0}}$$

Hence, at size α:

$$z_{\alpha/2} = k\sqrt{\frac{n}{\theta_0}} \Rightarrow k = z_{\alpha/2}\sqrt{\frac{\theta_0}{n}}$$

99. A point is to be selected from the unit interval $(0,1)$ randomly. Let $A_1 = (0, 1/4]$, $A_2 = (1/4,\ 1/2]$, $A_3 = (1/2,\ 3/4]$ and $A_4 = (3/4,\ \ 1)$. Random experiments are repeated independently for 80 times under the same conditions. Observed frequencies that these points fall into A_1, A_2, A_3 and A_4 are 6, 18, 20, and 36, respectively. Test

$$H_0\text{: The cdf is Beta}(2,1)$$

against

$$H_1\text{: The cdf is not Beta}(2,1)$$

at the 0.025 significance level ($\chi^2(0.025,3) = 9.3484$).

Solution. The standard method for testing Goodness of fit of a probability model is the Pearson's Chisq test which we will use for this problem. To streamline calculation, we tabulate the data as the following

bins	Observed Frequency (O_i)	Expected Frequency (E_i)
0.00~0.25	6	5
0.25~0.50	18	15
0.50~0.75	20	25
0.75~1.00	36	35

where the 'Expected Frequency' is calculated as $E_i = 80 \times \text{BetaProbability}[\text{bin}_i; 2,1]$. The Pearson's Chisq test employs the following test statistic

$$T = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \overset{\text{good fit}}{\sim} \chi^2(r - 1 - d)$$

which realizes here to $t = 1.829$, and $r = 4$ is the number of rows in the table, $d = 2$ as both parameters of the beta distribution are being tested/estimated. But one should note that the larger value of d the easier the rejection of null, so in the case that we are very comfortable with the null hypothesis, we may set $d = 0$, effectively reducing the probability of committing Type-1 error at the expense of a higher Type-2 error probability. We proceed with $d = 0$ for usability of the Hint to the problem. Since the critical value of $\chi^2(0.025,3) = 9.3484$ below with leads to acceptance of the null hypothesis, we will accept the null hypothesis at 0.025 significance level.

100. The following are the numbers of passengers carried on flights 136 and 137 between Chicago and Phoenix on 12 days :

232 and 189, 265 and 230, 249 and 236, 250 and 261,
255 and 249, 236 and 218, 270 and 258, 247 and 253,
249 and 251, 240 and 233, 257 and 254, 239 and 249.

Use the paired-sample sign test at the 0.05 level of significance to test the null hypothesis $H_0\colon \mu_1 = \mu_2$ (that on the average the two flights carry equally many passengers) against the alternative hypothesis $H_1\colon \mu_1 > \mu_2$ by calculating the exact p-value.

Solution. The paired sample sign test uses an array of paired data for which only the order in each pair is used. That is, the data has only 3 values: a '+' designates a pair if the first number is bigger; a '0' designates a pair if both numbers are equal; a '-' designates a pair if the first number is smaller. For the raw data here, they are transform as the 3-value vector:

$$+, +, +, -, +, +, +, -, -, +, +, -.$$

The length of the vector, or the sample size, is 12, the same as the number of pairs in the original raw data.

It is not difficult to understand the binomial distribution is being used as the probability vehicle for the paired sign test. If the null hypothesis is true, then one expects as many '+'s as '-'s, leading to the count of either of them following Binomial(n, 0.5) where $n = 12$ here is the sample size. We will just use the count of the '+'s, denoted by X, that is,

$$X \overset{H_0}{\sim} Binomial(n = 12, \quad p = 0.5)$$

The X is realized to $x = 8$. The p-value of $x = 8$ is

$$pvalue(x = 8) = 2 \times \left[\binom{12}{8} + \binom{12}{9} + \binom{12}{10} + \binom{12}{11} + \binom{12}{12}\right] \times 0.5^{12} = 0.3877$$

leading to the acceptance of $H_0\colon \mu_0 = \mu_1$ at 5% significance level. (The exact p-value simply means not to use normal approximation to binomial but directly calculate the probability from the binomial p.m.f.. The binomial coefficient embedded in the summation is deemed as tedious even for computers.)