Probability & Statistics (1)

Jointly Distributed Random Variables (II)

Asst. Prof. Chan, Chun-Hsiang

Master program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Applied Artificial Intelligence, Chung Yuan Christian University, Taoyuan, Taiwan

Outlines

1. Joint Distribution Functions
2. Independent Random Variables
3. Sums of Independent Random Variables
4. Conditional Distribution: Discrete Case
5. Conditional Distribution: Continuous Case
6. Order Statistics
7. Joint Probability Distribution of Functions of Random Variables
8. [#12] Assignment
9. Reference
10. Question Time

Sums of Independent Random Variables
Uniform Distribution

- 時常我們會需要用到兩個獨立的隨機變數相加的機率分配，譬如說現在有兩個隨機變數: X與Y。我們需要計算$X+Y$的$c.d.f.$就可以從它們各自的$p.d.f.(f_X \text{ and } f_Y)$計算得知:

$$F_{X+Y}(a) = P\{X+Y \le a\} = \iint_{x+y\le a} f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)dxf_Y(y)dy$$

$$= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy$$

F_{X+Y}的$c.d.f.$其實也就是F_X與F_Y的convolution。

Sums of Independent Random Variables
Uniform Distribution

Differentiation both side,

$$f_{X+Y}(a) = \frac{a}{da}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} \frac{d}{da}F_X(a-y)f_Y(y)dy$$

$$= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$$

Sums of Independent Random Variables
Uniform Distribution

- **範例一**

如果X與Y為獨立隨機變數且值域uniformly distributed on (0,1)，試問$X+Y$的$p.d.f.$。

Solution:

Let f_X and f_Y are the $p.d.f.$ of X and Y, respectively. Let $Z=X+Y$.

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx\,, where\ 0<x<1\ and\ 0<z-x<1$$

$$f_Z(z)=0, for\ z<0\ and\ z\geq 2; in\ other\ words, 0<z<2$$

Sums of Independent Random Variables
Uniform Distribution

(1) $\boldsymbol{Case\ 1: 0 < z \leq 1, f_X(x) f_Y(z-x)}$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

$To\ confirm\ f_Y(z-x) = 1, we\ need\ z - x \geq 0 \Rightarrow x \leq z$

$$\int_0^z 1 dx = z\,; f_Z(z) = z, for\ 0 < z \leq 1$$

(2) $\boldsymbol{Case\ 2: 1 < z < 2}$

$To\ confirm\ f_Y(z-x) = 1, we\ need\ z - x \leq 1 \Rightarrow x \geq z - 1$

$$\int_{z-1}^1 1 dx = 2 - z\,; f_Z(z) = 2 - z, for\ 1 < z < 2$$

Sums of Independent Random Variables
Uniform Distribution

As a result,

$$f_Z(z) = f_{X+Y}(z) = \begin{cases} z, 0 \leq z \leq 1 \\ 2 - z, 1 < z < 2 \\ 0, otherwise \end{cases}$$

Because of the shape of $p.d.f.$, the random variable of $X + Y$ is a $triangular\ distribution$.

Sums of Independent Random Variables
Uniform Distribution

假設$X_1, X_2, \ldots, X_n$為independent uniform random variable，使得

$$F_n(x) = P\{X_1 + \cdots + X_n \leq x\}$$

雖然$F_n(x)$的general form非常難表示，不過$x \leq 1$的時候，就可以用mathematical induction得到

$$F_n(x) = \frac{x^n}{n!}, 0 \leq x \leq 1$$

已知$n = 1$是正確的，假設

$$F_{n-1}(x) = \frac{x^{n-1}}{(n-1)!}, 0 \leq x \leq 1$$

Sums of Independent Random Variables
Uniform Distribution

於是就可以表示成

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n, X_i \; is \; nonnegative, where \; 0 \le x \le 1$$

$$F_n(x) = \int_0^1 F_{n-1}(x-y) f_{X_n}(y)dy = \frac{1}{(n-1)!}\int_0^x (x-y)^{n-1}dy = \frac{x^n}{n!}$$

$for \; n = 1, X \; follows \; a \; uniform \; distribution$

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

$for \; n = 2, X \; follows \; a \; triangular \; distribution$

$$f_X(x) = \begin{cases} x & 0 \le x \le 1 \\ 2 - x & 1 \le x \le 2 \end{cases}$$

$for \; n = 3,$

$$f_X(x) = \begin{cases} \frac{1}{2}x^2 & 0 \le x \le 1 \\ \frac{1}{2}(-2x^2 + 6x - 3) & 1 \le x \le 2 \\ \frac{1}{2}(3-x)^2 & 2 \le x \le 3 \end{cases}$$

Sums of Independent Random Variables
Uniform Distribution

$for\ n = 4,$

$$f_X(x) = \begin{cases} \frac{1}{6}x^3 & 0 \le x \le 1 \\ \frac{1}{6}(-3x^3 + 12x^2 - 12x + 4) & 1 \le x \le 2 \\ \frac{1}{6}(3x^3 - 24x^2 + 60x - 44) & 2 \le x \le 3 \\ \frac{1}{6}(4 - x)^3 & 3 \le x \le 4 \end{cases}$$

Sums of Independent Random Variables
Uniform Distribution

$for\ n = 5,$

$$f_X(x) = \begin{cases} \frac{1}{24}x^4 & 0 \le x \le 1 \\ \frac{1}{24}(-4x^4 + 20x^3 - 30x^2 + 20x - 5) & 1 \le x \le 2 \\ \frac{1}{24}(6x^4 - 60x^3 + 210x^2 - 300x + 155) & 2 \le x \le 3 \\ \frac{1}{24}(-4x^4 + 60x^3 - 330x^2 + 780x - 655)^3 & 3 \le x \le 4 \\ \frac{1}{24}(5 - x)^4 & 4 \le x \le 5 \end{cases}$$

Sums of Independent Random Variables
Normal Distribution

Proposition 1

If X_i, $i = 1, \ldots, n$, are independent random variables that are normally distributed with respective parameters μ_i, $\sigma_i^2, i = 1, \ldots, n$, then $\sum_{i=1}^{n} X_i$ is normally distributed with parameters $\sum_{i=1}^{n} \mu_i$ and $\sum_{i=1}^{n} \sigma_i^2$.

Proof:

Let X and Y be independent normal random variables with X having mean 0 and variance σ^2 and Y having mean 0 and variance 1. We will determine the density function of $X + Y$ by utilizing $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$.

Sums of Independent Random Variables
Normal Distribution

For independent random variables X and Y, the distribution f_Z of $Z = X + Y$ equals the convolution of f_X and f_Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx$$

Given that f_X and f_Y are normal densities,

$$f_X(x) = \mathcal{N}(x; \mu_X, \sigma_X^2) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right\}$$

$$f_Y(y) = \mathcal{N}(y; \mu_Y, \sigma_Y^2) = \frac{1}{\sigma_Y \sqrt{2\pi}} \exp\left\{-\frac{(y - \mu_Y)^2}{2\sigma_Y^2}\right\}$$

Sums of Independent Random Variables
Normal Distribution

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}\right] \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma_X\sqrt{2\pi}\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{\sigma_X^2(z-x-\mu_Y)^2 + \sigma_Y^2(x-\mu_X)^2}{2\sigma_X^2\sigma_Y^2}\right] dx$$

$$= \int_{\infty}^{\infty} \frac{1}{\sigma_X\sqrt{2\pi}\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{\sigma_X^2(z^2+x^2+\mu_Y^2-2xz-2z\mu_Y+2x\mu_Y) \quad + \sigma_Y^2(x^2+\mu_X^2-2x\mu_X)}{2\sigma_X^2\sigma_Y^2}\right] dx$$

$$= \int_{\infty}^{\infty} \frac{1}{\sigma_X\sqrt{2\pi}\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{x^2(\sigma_X^2+\sigma_Y^2) - 2x(\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X) + \sigma_X^2(z^2+\mu_Y^2-2z\mu_Y) + \sigma_Y^2\mu_X^2}{2\sigma_X^2\sigma_Y^2}\right] dx$$

Let $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}$

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sigma_Z\sqrt{2\pi}} \frac{1}{\frac{\sigma_X\sigma_Y}{\sigma_Z}\sqrt{2\pi}} \exp\left[-\frac{\frac{x^2(\sigma_Z^2)}{\sigma_Z^2} - 2x\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2} + \frac{\sigma_X^2(z^2+\mu_Y^2-2z\mu_Y)+\sigma_Y^2\mu_X^2}{\sigma_Z^2}}{2\left(\frac{\sigma_X\sigma_Y}{\sigma_Z}\right)^2}\right] dx$$

Sums of Independent Random Variables
Normal Distribution

$$f_Z(z)=\int_{-\infty}^{\infty}\frac{1}{\sigma_Z\sqrt{2\pi}}\frac{1}{\frac{\sigma_X\sigma_Y}{\sigma_Z}\sqrt{2\pi}}\exp\left[-\frac{x^2-2x\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2}+\frac{\sigma_X^2(z^2+\mu_Y^2-2z\mu_Y)+\sigma_Y^2\mu_X^2}{\sigma_Z^2}}{2\left(\frac{\sigma_X\sigma_Y}{\sigma_Z}\right)^2}\right]dx$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sigma_Z\sqrt{2\pi}}\frac{1}{\frac{\sigma_X\sigma_Y}{\sigma_Z}\sqrt{2\pi}}\exp\left[-\frac{\left(x-\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2}\right)^2-\left(\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2}\right)^2+\frac{\sigma_X^2(z^2+\mu_Y^2-2z\mu_Y)+\sigma_Y^2\mu_X^2}{\sigma_Z^2}}{2\left(\frac{\sigma_X\sigma_Y}{\sigma_Z}\right)^2}\right]dx$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sigma_Z\sqrt{2\pi}}\exp\left[-\frac{-(\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X)^2+\sigma_Z^2(\sigma_X^2(z^2+\mu_Y^2-2z\mu_Y)+\sigma_Y^2\mu_X^2)}{2\sigma_Z^2(\sigma_X\sigma_Y)^2}\right]\frac{1}{\frac{\sigma_X\sigma_Y}{\sigma_Z}\sqrt{2\pi}}\exp\left[\frac{\left(x-\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sigma_Z}\right)^2}\right]dx$$

$$=\frac{1}{\sigma_Z\sqrt{2\pi}}\exp\left[-\frac{(z-(\mu_X+\mu_Y))^2}{2\sigma_Z^2}\right]\int_{-\infty}^{\infty}\frac{1}{\frac{\sigma_X\sigma_Y}{\sigma_Z}\sqrt{2\pi}}\exp\left[\frac{\left(x-\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sigma_Z}\right)^2}\right]dx$$

Sums of Independent Random Variables
Normal Distribution

$$f_Z(z) = \frac{1}{\sigma_Z\sqrt{2\pi}} \exp\left[-\frac{\left(z-(\mu_X+\mu_Y)\right)^2}{2\sigma_Z^2}\right] \int_{-\infty}^{\infty} \frac{1}{\frac{\sigma_X\sigma_Y}{\sigma_Z}\sqrt{2\pi}} \exp\left[\frac{\left(x - \frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_Z^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sigma_Z}\right)^2}\right] dx$$

The expression in the integral is a normal density distribution on x, and so the integral evaluates to 1. Therefore, …

$$f_Z(z) = \frac{1}{\sigma_Z\sqrt{2\pi}} \exp\left[-\frac{\left(z-(\mu_X+\mu_Y)\right)^2}{2\sigma_Z^2}\right]$$

Sums of Independent Random Variables
Normal Distribution

- **範例二**

假設籃球比賽一季有44場比賽: 26場與一級球隊比賽；18場與二級球隊比賽。假設與一級球隊比賽的勝率為0.4，而與二級球隊比賽的勝率為0.7，而每一場比賽接為獨立事件，則:

(1) 某一球隊贏25場的機率為何?

(2) 某一球隊贏一級球隊多於二級球隊的機率為何?

Sums of Independent Random Variables
Normal Distribution

Solution:

(1)某一球隊贏25場的機率為何?

令X_A與X_B分別為贏一級球隊與二級球隊的比賽場數，且X_A與X_B為independent binomial random variable。

$$E[X_A]=26\times0.4=10.4; Var(X_A)=26\times0.4\times0.6=6.24$$
$$E[X_B]=18\times0.7=12.6; Var(X_B)=18\times0.7\times0.3=3.78$$

此時，我們可以用normal approximation來逼近binomial，X_A與X_B會逼近normal distribution相同參數(μ,σ^2)的分布。

依據**Proposition 1**可以得知，X_A+X_B所逼近的常態分配參數為$(\mu=10.4+12.6=23,\sigma^2=6.24+3.78=10.02)$

Sums of Independent Random Variables
Normal Distribution

$$P\{X_A + X_B \geq 25\} = P\{X_A + X_B \geq 24.5\}$$

$$= P\left\{\frac{X_A + X_B - 23}{\sqrt{10.02}} \geq \frac{24.5 - 23}{\sqrt{10.02}}\right\} \approx P\left\{Z \geq \frac{1.5}{\sqrt{10.02}}\right\}$$

$$\approx 1 - P\{Z < 0.4739\} \approx 0.3178$$

Sums of Independent Random Variables
Normal Distribution

(2)某一球隊贏一級球隊多於二級球隊的機率為何?

此提要求的就是 $P\{X_A - X_B \geq 1\}$，對於 $X_A - X_B$ 來說，其 normal distribution 的參數為 $(\mu = 10.4 - 12.6 = -2.2, \sigma^2 = 6.24 + 3.78 = 10.02)$

$$\because Var(X - Y) = Var\big(X + (-Y)\big) = Var(X) + Var(-Y),$$

$$since\ Var(-Y) = Var(Y)$$

$$\therefore Var(X - Y) = Var(X) + Var(Y)$$

$$P\{X_A - X_B \geq 1\} = P\{X_A + X_B \geq 0.5\}$$

$$= P\left\{\frac{X_A - X_B - (-2.2)}{\sqrt{10.02}} \geq \frac{0.5 - (-2.2)}{\sqrt{10.02}}\right\} \approx P\left\{Z \geq \frac{2.7}{\sqrt{10.02}}\right\}$$

$$\approx 1 - P\{Z < 0.8530\} \approx 0.1968$$

Sums of Independent Random Variables
Lognormal Distribution

如果$\log(Y)$為一個normal random variable (μ, σ^2)，則Y就可以被定義為一個lognormal random variable (μ, σ)。Y可以被表示為

$$Y = e^X$$

where X is a normal random variable

Sums of Independent Random Variables
Normal Distribution

- **範例三**

從某個固定時間開始，令$S(n)$表示在 n 週結束時某種證券的價格, $n \geq 1$。常見的價格預測方式為價格比$S(n)/S(n-1)$, $n \geq 1$，為 independent and identical distributed lognormal random variables。假設其$\mu = 0.0165$，$\sigma = 0.0730$，是求以下的機率:

(1) 證券價格在未來兩周內都會增長

(2) 兩週後的價格高於今天的價格

Sums of Independent Random Variables
Normal Distribution

Solution:

(1) 證券價格在未來兩周內都會增長

令Z為standard normal random variable，為了解第一題，我們可以利用題目的線索，也就是持續增長，亦即$\log(x)$必須隨著x增加而增加，所以x必須大於1，也可以表示為$\log(x) > \log(1) \Rightarrow \log(x) - \log(1) > 0$。

$$P\left\{\frac{S(1)}{S(0)} > 1\right\} = P\left\{\log\left(\frac{S(1)}{S(0)}\right) > 0\right\} = P\left\{Z > \frac{0 - 0.0165}{0.0730}\right\}$$

$$= P\{Z < 0.2260\} = 0.5894; \textit{each of next two weeks} = (0.5894)^2$$

$$= 0.3474$$

Sums of Independent Random Variables
Normal Distribution

(2)兩週後的價格高於今天的價格

$$P\left\{\frac{S(2)}{S(0)} > 1\right\} = P\left\{\frac{S(2)}{S(1)}\frac{S(1)}{S(0)} > 1\right\} = P\left\{\log\left(\frac{S(2)}{S(1)}\right) + \log\left(\frac{S(1)}{S(0)}\right) > 0\right\}$$

這就可以視為兩個independent normal random variable相加，所以相加後的normal random variable ($\mu = 0.0165 + 0.0165 = 0.0330, \sigma^2 = 2 \times (0.0730)^2$)

$$P\left\{\frac{S(2)}{S(0)} > 1\right\} = P\left\{Z > \frac{0 - 0.0330}{0.0730\sqrt{2}}\right\} = P\{Z < 0.31965\} = 0.6254$$

Sums of Independent Random Variables
Poisson and Binomial Distribution

- **範例四**

如果X與Y為independent Poisson random variable，其參數分別為λ_1與λ_2，是求出$X+Y$的$p.m.f.$:

Solution:

我們可以假設$\{X+Y=n\}$，也可以拆成$\{X=k, Y=n-k\}, 0 \le k \le n$，則

$$P\{X+Y=n\} = \sum_{k=0}^{n} P\{X=k, Y=n-k\} = \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\}$$

$$= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!} e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} = e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!\,(n-k)!}$$

Sums of Independent Random Variables
Poisson and Binomial Distribution

$$= \sum_{k=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} = e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k \lambda_2^{n-k}}{k!\,(n-k)!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \lambda_1^k \lambda_2^{n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n$$

故$X + Y$為一個Poisson distribution $(\lambda_1 + \lambda_2)$

Sums of Independent Random Variables
Poisson and Binomial Distribution

- 範例五

如果X與Y為independent binomial random variables，其參數分別為(n,p)與(m,p)，試問$X+Y$的$p.m.f.$為何?

Solution:

$X+Y$的binomial random variable的參數為$(n+m,p)$，故:

$$P\{X+Y\}=\sum_{i=0}^{n} P\{X=i,Y=k-i\}=\sum_{i=0}^{n} P\{X=i\}P\{Y=k-i\}$$

$$=\sum_{i=0}^{n}\binom{n}{i}p^{i}q^{n-i}\binom{m}{k-i}p^{k-i}q^{m-k+i},$$

Sums of Independent Random Variables
Poisson and Binomial Distribution

$$P\{X+Y\} = \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-k+i}$$

$$= p^{i+k-i} q^{n-i+m-k+i} \sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i} = p^k q^{n+m-k} \sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i}$$

$$\Rightarrow P\{X+Y\} = p^k q^{n+m-k} \sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i}$$

$since\ P\{X+Y\} \sim binomial(n+m,k), thus \Rightarrow$

$$\sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k}$$

Sums of Independent Random Variables
Geometric Distribution

- **範例六**

令$X_1, \dots, X_n$為independent geometric random variables，且其機率為$p_i\ for\ i = 1, \dots, n$。試問$S_n = \sum_{i=1}^{n} X_i$的$p.m.f.$為何?

Solution:

我們可以試想一個投擲硬幣的情境，X_n就可以想做有n枚硬幣，針對第i枚硬幣出現正面的機率為p_i，於是乎$S_n = \sum_{i=1}^{n} X_i$就可以被當作投擲了n次正面的機率，所以就可以被視為negative binomial random variable。

$$P\{S_n = k\} = \binom{k-1}{n-1} p^n (1-p)^{k-n}, k \geq n$$

Sums of Independent Random Variables
Geometric Distribution

$$P\{S_n = k\} = \binom{k-1}{n-1} p^n (1-p)^{k-n}, k \geq n$$

為了歸納其S_n 的$p.m.f.$，假設所有的p_i都不一樣。我們先考慮$n = 2$的情形，令$q_j = 1 - p_j, j = 1,2$，則

$$P\{S_2 = k\} = \sum_{j=1}^{k-1} P\{X_1 = j, X_2 = k - j\} = \sum_{j=1}^{k-1} P\{X_1 = j\} P\{X_2 = k - j\}$$

$$= \sum_{j=1}^{k-1} p_1 q_1^{j-1} p_2 q_2^{k-j-1} = p_1 p_2 q_2^{k-2} \sum_{j=1}^{k-1} \left(\frac{q_1}{q_2}\right)^{j-1}$$

$$= p_1 p_2 q_2^{k-2} \frac{1 - (q_1/q_2)^{k-1}}{1 - q_1/q_2} = p_1 p_2 q_2^{k-2} q_2 \frac{1 - (q_1/q_2)^{k-1}}{q_2 - q_1}$$

Sums of Independent Random Variables
Geometric Distribution

$$P\{S_2 = k\} = p_1 p_2 q_2^{k-2} q_2 \frac{1-(q_1/q_2)^{k-1}}{q_2 - q_1} = p_1 p_2 q_2^{k-1} \frac{1-(q_1/q_2)^{k-1}}{q_2 - q_1}$$

$$= \frac{p_1 p_2 q_2^{k-1}}{q_2 - q_1} - \frac{p_1 p_2 q_2^{k-1} \left(\frac{q_1}{q_2}\right)^{k-1}}{q_2 - q_1} = \frac{p_1 p_2 q_2^{k-1}}{q_2 - q_1} - \frac{p_1 p_2 q_1^{k-1}}{q_2 - q_1}$$

$$= p_2 q_2^{k-1} \frac{p_1}{q_2 - q_1} - p_1 q_1^{k-1} \frac{p_2}{q_2 - q_1}$$

$$= p_2 q_2^{k-1} \frac{p_1}{p_1 - p_2} - p_1 q_1^{k-1} \frac{p_2}{p_1 - p_2}$$

$$= p_2 q_2^{k-1} \frac{p_1}{p_1 - p_2} + p_1 q_1^{k-1} \frac{p_2}{p_2 - p_1}$$

Sums of Independent Random Variables
Geometric Distribution

那如果是$n=3$的時候，$P\{S_3=k\}$，

$$P\{S_3=k\}=\sum_{j=1}^{k-1}P\{S_2=j,X_3=k-j\}=\sum_{j=1}^{k-1}P\{S_2=j\}P\{X_3=k-j\}$$

$$=p_1q_1^{k-1}\frac{p_2}{p_2-p_1}\frac{p_3}{p_3-p_1}+p_2q_2^{k-1}\frac{p_1}{p_1-p_2}\frac{p_3}{p_3-p_2}+p_3q_3^{k-1}\frac{p_1}{p_1-p_3}\frac{p_2}{p_2-p_3}$$

Sums of Independent Random Variables
Geometric Distribution

根據剛剛的規律，我們可以得出一個新的proposition。

- **Proposition 2**

Let $X_1, \ldots, X_n$ 為 independent geometric random variables, with X_i having parameters p_i, for $i = 1, \ldots, n$. If all the p_i are distinct, then , for $k \geq n$,

$$P\{S_n = k\} = \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i}$$

Conditional Distributions: Discrete Case

- 給定任何兩個事件E與F，則F成立前提下E所發生的條件機率(conditional probability)為:

$$P(E|F) = \frac{P(EF)}{P(F)}, where\ P(F) > 0$$

- 當今天有兩個discrete random variable X與Y，在Y成立前提下X發生的$conditional\ probability\ mass\ function$為

$$p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}} = \frac{p(x,y)}{p_Y(y)},$$

$$where\ p_Y > 0$$

Conditional Distributions: Discrete Case

- 那計算在Y成立前提下X發生的$cumulative\ conditional\ p.m.f.$為

$$F_{X|Y}(x|y) = P\{X \le x, Y \le y\} = \sum_{a \le x} p_{X|Y}(a|y)$$

- 如果X與Y相互獨立的話，

$$p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}} = \frac{P\{X = x\}P\{Y = y\}}{P\{Y = y\}}$$

$$= P\{X = x\}$$

Conditional Distributions: Discrete Case

- **範例七**

令$p(x,y)$為X與Y的$joint\ p.m.f.$，

$$p(0,0)=0.4; p(0,1)=0.2; p(1,0)=0.1; p(1,1)=0.3$$

請問$Y=1$的前提下X發生的$conditional\ p.m.f.$為何?

Solution:

$$p_Y(1)=\sum_x p(x,1)=p(0,1)+p(1,1)=0.2+0.3=0.5$$

$$p_{X|Y}(0|1)=\frac{p(0,1)}{p_Y(1)}=\frac{2}{5};\ p_{X|Y}(1|1)=\frac{p(1,1)}{p_Y(1)}=\frac{3}{5}$$

Conditional Distributions: Discrete Case

- **範例八**

假設X與Y皆為independent Poisson random variables，其參數分別為λ_1與λ_2，試問在$X+Y=n$的前提下X的conditional probability為何?

Solution:

本題要計算$X+Y=n$前提下X的conditional probability mass function。

$$P\{X=k \mid X+Y=n\}=\frac{P\{X=k, X+Y=n\}}{P\{X+Y=n\}}=\frac{P\{X=k, Y=n-k\}}{P\{X+Y=n\}}$$

$$=\frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}=\frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1}$$

Conditional Distributions: Discrete Case

$$P\{X = k \mid X + Y = n\} = \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1}$$

$$= \frac{n!}{(n-k)!\,k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k}$$

The conditional distribution of X given that $X + Y = n$ is the binomial distribution with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$

Conditional Distributions: Continuous Case

- 如果X與Y的joint probability density function為$f(x,y)$，則在給定$Y=y$的前提下X的conditional probability density function可以被定義，對於每一個y其$f_Y(y)>0$，

$$f_{X+Y}(x|y)dx=\frac{f(x,y)dxdy}{f_Y(y)dy}\approx\frac{P\{x\le X\le x+dx,y\le Y\le y+dy\}}{P\{y\le y\le y+dy\}}$$

$$=P\{x\le X\le x+dx|y\le Y\le y+dy\}$$

- 如果X與Y是jointly continuous random variable，則對於任何集合A

$$P\{X\in A|Y=y\}=\int_A f_{X|Y}(x|y)dx$$

Conditional Distributions: Continuous Case

- Let $A = (-\infty, a]$，則在給定$Y = y$前提下X的conditional cumulative distribution function為

$$F_{X+Y}(a|y) \equiv P\{X \leq a|Y = y\} = \int_{-\infty}^{a} f_{X+Y}(x|y)dx$$

Conditional Distributions: Continuous Case

- 範例九

X與Y的joint density function為

$$f(x,y)=\begin{cases}\frac{12}{5}x(2-x-y) & 0<x<1, 0<y<1\\ 0 & otherwise\end{cases}$$

計算$Y=y$前提下X的conditional density為何?

Solution:

$$f_{X+Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)dx}=\frac{x(2-x-y)}{\int_0^1 x(2-x-y)dx}=\frac{x(2-x-y)}{\frac{2}{3}-\frac{y}{2}}$$

$$=\frac{6x(2-x-y)}{4-3y}$$

Conditional Distributions: Continuous Case

- **範例十**

假設X與Y的joint density function為

$$f(x,y)=\begin{cases}\dfrac{e^{-\frac{x}{y}}e^{-y}}{y} & 0<x<\infty, 0<y<\infty\\ 0 & otherwise\end{cases}$$

試問$P\{X>1|Y=y\}$。

Solution:

$$f_{X+Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{\dfrac{e^{-\frac{x}{y}}e^{-y}}{y}}{e^{-y}\int_0^{\infty}\frac{1}{y}e^{-\frac{x}{y}}dx}=\frac{1}{y}e^{-\frac{x}{y}}$$

Conditional Distributions: Continuous Case

$$f_{X+Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{e^{-\frac{x}{y}}e^{-y}}{y}}{e^{-y}\int_0^\infty \frac{1}{y}e^{-\frac{x}{y}}dx} = \frac{1}{y}e^{-\frac{x}{y}}$$

Hence,

$$P\{X > 1|Y = y\} = \int_1^\infty \frac{1}{y}e^{-\frac{x}{y}}dx = -e^{-\frac{x}{y}}\Big|_1^\infty = e^{-\frac{1}{y}}$$

Conditional Distributions: Continuous Case

If X and Y are independent continuous random variables, the conditional density of X given that $Y = y$ is just the unconditional density of X. This is so because, in the independent case,

$$f_{X+Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

Order Statistics

- Let $X_1, X_2, \dots, X_n$ be n independent and identically distributed continuous random variables having a common density f and distribution function F. Define

$$X_{(1)} = smallest\ of\ X_1, X_2, \dots, X_n$$
$$X_{(2)} = second\ smallest\ of\ X_1, X_2, \dots, X_n$$
$$\vdots$$
$$X_{(j)} = jth\ smallest\ of\ X_1, X_2, \dots, X_n$$
$$\vdots$$
$$X_{(n)} = largest\ of\ X_1, X_2, \dots, X_n$$

Order Statistics

- The ordered values $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are known as the order statistics corresponding to the random variables $X_1, X_2, \ldots, X_n$. In other words, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the ordered values of $X_1, X_2, \ldots, X_n$.
- The joint density function of the ordered statistics is obtained by noting that the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ will take on the values $x_1 \leq x_2 \leq \cdots \leq x_n$ if and only if, for some permutation $(i_1, i_2, \ldots, i_n)$ of $(1,2, \ldots, n)$,

$$X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}$$

Order Statistics

- Since, for any permutation $(i_1, i_2, \dots, i_n)$ of $(1,2,\dots,n)$,

$$P\left\{x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2}, \dots, x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2}\right\}$$

$$\approx \varepsilon^n f_{X_1,\dots X_n}\left(x_{i_1}, \dots x_{i_n}\right) = \varepsilon^n f\left(x_{i_1}\right) \dots f\left(x_{i_n}\right) = \varepsilon^n f(x_i) \dots f(x_n)$$

It follows that, for $x_1 < x_2 < \cdots < x_n$,

$$P\left\{x_1 - \frac{\varepsilon}{2} < X_{(1)} < x_1 + \frac{\varepsilon}{2}, \dots, x_n - \frac{\varepsilon}{2} < X_{(n)} < x_n + \frac{\varepsilon}{2}\right\}$$

$$\approx n!\, \varepsilon^n f(x_1) \dots f(x_n)$$

Dividing by ε^n and letting $\varepsilon \to 0$ yields

$$\varepsilon^n f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots x_n) = n!\, \varepsilon^n f(x_1) \dots f(x_n)$$

$$f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots x_n) = n!\, f(x_1) \dots f(x_n), x_1 < x_2 < \cdots < x_n$$

Order Statistics

- **範例十一**

有3個人隨機位在(distributed at random)一條1公里的道路上，試問: 不存在任兩人相距小於d公里($d < \frac{1}{2}$)的機率。

Solution:

distributed at random這句話可以當作independent and uniformly distributed。令X_i為第i個人在這條道路上的位置，我們要求的機率就可以表示為$P\{X_{(i)} > X_{(i-1)} + d, i = 2,3\}$

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3) = 3!, 0 < x_1 < x_2 < x_3 < 1$$

Order Statistics

$$P\{X_{(i)} > X_{(i-1)} + d, i = 2,3\} = \iiint_{x_i > x_{j-1}+d} f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3)\, dx_1 dx_2 dx_3$$

$$= 3! \int_0^{1-2d} \int_{x_1+d}^{1-d} \int_{x_2+d}^{1} dx_3 dx_2 dx_1 = 6 \int_0^{1-2d} \int_{x_1+d}^{1-d} (1 - x_2 - d) dx_2 dx_1$$

$$let\ (1 - x_2 - d) = y_2$$

$$= 6 \int_0^{1-2d} \int_0^{1-d-(x_1+d)=1-2d-x_1} y_2 dy_2 dx_1 = 3 \int_0^{1-2d} (1 - 2d - x_1)^2 dx_1$$

$$= 3 \int_0^{1-2d} y_1^2 dy_1 = (1 - 2d)^3 \Rightarrow if\ there\ are\ n\ people, \ldots [1 - (n-1)d]^n, where\ d < \frac{1}{n-1}$$

Order Statistics

- The density function of the $j-th$-order statistics $X_{(j)}$ can be obtained either by integrating the joint density function (pp.49) or by direct reasoning as follows: In order for $X_{(j)}$ to equal x, it is necessary for $j-1$ of the n values $X_1, \ldots, X_n$ to be less than $x, n-j$ of them to be greater than x, and 1 of them to equal x.
- Now, the probability density that any given set of $j-1$ of the X_i's are less than x, another given set of $n-j$ are all greater than x, and the remaining value is equal to x equals.

$$[F(x)]^{j-1}[1-F(x)^{n-j}]f(x)$$

Therefore,

$$\binom{n}{j-1, n-j, 1} = \frac{n!}{(n-j)!\,(j-1)!}$$

Order Statistics

independent

$$F_n(y_n) = P\{Y_n \le y_n\} = P\{Y_1 \le y_n, Y_2 \le y_n, \dots, Y_n \le y_n\} = [F(y_n)]^n$$

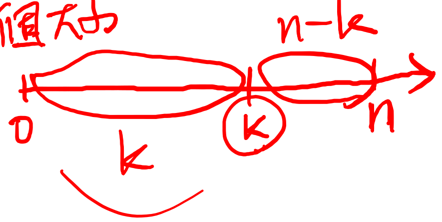
differentiation

$$f_n(y_n) = \frac{\partial F_n(y_n)}{\partial y} = n[F(y_n)]^n f(y_n)$$ ；令 $y = y_n$ 某一個常數 $\ge Y_1 \sim Y_n$

$$F_k(y) = P\{Y_k < y\} = \sum_{i=k}^{n} \binom{n}{i} (F(y))^i (1-F(y))^{n-i} = \sum_{i=k}^{n-1} \binom{n}{i} (F(y))^i (1-F(y))^{n-i} + (F(y))^n$$

$i=n$: $\binom{n}{n}(F(y))^n (1-F(y))^{n-n=0}$

第k個大的



Differentiation

$$f_k(y) = \frac{\partial F_k(y)}{\partial y} = \sum_{i=k}^{n-1} \binom{n}{i} i (F(y))^{i-1} f(y) (1-F(y))^{n-i} - \sum_{i=k}^{n-1} \binom{n}{i} (F(y))^i (n-i)(1-F(y))^{n-i-1} f(y) + n(F(y))^{n-1} f(y)$$

$$\Rightarrow \sum_{i=k}^{n-1} \frac{n!}{(i-1)!(n-i)!} (F(y))^{i-1} f(y)(1-F(y))^{n-i} - \sum_{i=k}^{n-1} \frac{n!}{i!(n-i-1)!} (F(y))^i (1-F(y))^{n-i-1} f(y) + n(F(y))^{n-1} f(y)$$

Order Statistics

$$\hookrightarrow \sum_{i=k}^{n-1} \frac{n!}{(i-1)!\,(n-i)!}(F(y))^{i-1}\cdot f(y)\,[1-F(y)]^{n-i} - \sum_{i=k}^{n-1}\frac{n!}{i!\,(n-i-1)!}(F(y))^{i}\,(1-F(y))^{n-i-1}\cdot f(y) + n(F(n))^{n-1} f(y)$$

if $i=k$: $$\frac{n!}{(k-1)!\,(n-k)!}(F(y))^{k-1}\cdot f(y)\,[1-F(y)]^{n-k} - \frac{n!}{k!\,(n-k-1)!}(F(y))^{k}\,(1-F(y))^{n-k-1}\cdot f(y) + n(F(n))^{n-1} f(y)$$

$i=k+1$: $$\frac{n!}{k!\,(n-k-1)!}(F(y))^{k}\cdot f(y)\,[1-F(y)]^{n-k-1} - \frac{n!}{(k+1)!\,(n-k-2)!}(F(y))^{k+1}\,(1-F(y))^{n-k-2}\cdot f(y) + n(F(n))^{n-1} f(y)$$

⋮

$i=n-1$: $$\frac{n!}{(n-2)!\,(n-n+1)!}(F(y))^{n-2}\cdot f(y)\,[1-F(y)]^{n-n+1} - \frac{n!}{(n-1)!\,(n-(n-1)-1)!}(F(y))^{n-1}\,(1-F(y))^{n-(n-1)-1}\cdot f(y)$$

$$\frac{n!}{(k-1)!\,(n-k)!}(F(y))^{k-1}\cdot f(y)[1-F(y)]^{n-k} - n\cdot(F(y))^{n-1}\cdot f(y) + n(F(n))^{n-1}\cdot f(y)$$

$$\rightarrow \frac{n!}{(k-1)!\,(n-k)!}(F(y))^{k-1}\cdot f(y)\,[1-F(y)]^{n-k}$$

Order Statistics

- Different partitions of the n random variables $X_1, \ldots X_n$ into the preceding three groups, it follows that the density function of $X_{(j)}$ is given by

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!\,(j-1)!}[F(x)]^{j-1}[1 - F(x)^{n-j}]f(x)$$

Order Statistics

- **範例十二**

如果今天有一個樣本含有$2n+1$個random variables ($2n+1$ independent and identically distributed random variables)，則第$n+1$小的稱為sample median。假設今天樣本大小為3，且為uniform distribution over $(0,1)$，試問: sample median落在$\frac{1}{4}$與$\frac{3}{4}$之間的機率為何?

Order Statistics

Solution:

根據pp.51所導出來的公式得知$X_{(2)}$

$$f_{X_{(2)}}(x)=\frac{3!}{1!\,1!}x(1-x), 0<x<1$$

$Hence,$

$$P\left\{\frac{1}{4}<X_{(2)}<\frac{3}{4}\right\}=6\int_{1/4}^{3/4}x(1-x)\,dx=6\left\{\frac{x^2}{2}-\frac{x^3}{3}\right\}\Bigg|_{x=1/4}^{x=3/4}=\frac{11}{16}$$

Order Statistics

$The\ c.d.f.\ of\ X_{(j)}\ can\ be\ found\ by\ integrating\ pp.51\ formula$

$$F_{X_{(j)}}(y) = \frac{n!}{(n-j)!\,(j-1)!}\int_{-\infty}^{y}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)dx$$

$$F_{X_{(j)}}(y) = P\{X_{(j)} \le y\} = P\{j\ or\ more\ of\ the\ X_i\ 's\ are \le y\}$$

$$= \sum_{k=j}^{n}\binom{n}{k}[F(y)]^k[1-F(y)]^{n-k}$$

$We\ take\ F\ to\ be\ the\ uniform\ (0,1)\ distribution\ [that\ is, f(x) = 1, 0 < x < 1],$
$then\ we\ obtain\ the\ interesting\ analytical\ identity$

$$\sum_{k=j}^{n}\binom{n}{k}y^k(1-y)^{n-k} = \frac{n!}{(n-j)!\,(j-1)!}\int_{0}^{y}x^{j-1}(1-x)^{n-j}dx\ , where\ 0 \le y \le 1$$

Order Statistics

Referring to *pp.51*

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!\,(j-1)!}[F(x)]^{j-1}[1-F(x)^{n-j}]f(x)$$

We can obtain…

$$f_{X_{(i)},X_{(j)}}\left(x_i, x_j\right)$$
$$= \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}[F(x_i)]^{i-1} \times \left[F\left(x_j\right) - F(x_i)\right]^{j-i-1}\left[1-F\left(x_j\right)\right]^{n-j} f(x_i) f\left(x_j\right)$$
$$for\ all\ x_i < x_j$$

Joint Probability Distribution of Functions of Random Variables

- Let X_1 and X_2 be jointly continuous random variables with joint probability density function $f_{X1,X2}$. It is sometimes necessary to obtain the joint distribution of the random variables Y_1 and Y_2, which arise as functions of X_1 and X_2. Specifically, suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2.

Joint Probability Distribution of Functions of Random Variables

- Assume that the functions g_1 and g_2 satisfy the following conditions:
  1. The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, with solutions given by, say, $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$.
  2. The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that the 2×2 determinant.

Joint Probability Distribution of Functions of Random Variables

$$J(x_1,x_2)=\begin{vmatrix}\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2}\\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2}\end{vmatrix}\equiv\frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2}-\frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1}\neq 0, at\ all\ point\ (x_1,x_2)$$

- Under these two conditions, it can be shown that the random variables Y_1 and Y_2 are jointly continuous with joint density function given by $f_{Y_1Y_2}(y_1,y_2)=f_{X_1,X_2}(x_1,x_2)|J(x_1,x_2)|^{-1}, where\ x_1=h_1(y_1,y_2), x_2=h_2(y_1,y_2)$
- A proof of above equation would proceed along the following lines

$$P\{Y_1\le y_1,Y_2\le y_2\}=\iint_{\substack{(x_1,x_2):\\ g_1(x_1,x_2)\le y_1\\ g_2(x_1,x_2)\le y_2}} f_{X_1,X_2}(x_1,x_2)dx_1dx_2$$

Joint Probability Distribution of Functions of Random Variables

- 範例十三

Let X_1 and X_2 be joint continuous random variables with probability density function f_{X_1,X_2}. Let $Y_1 = X_1 + X_2,\ \ Y_2 = X_1 - X_2$. Find the joint density function of Y_1 and Y_2 in terms of f_{X_1,X_2}.

Solution:

$$Let\ g_1(x_1, x_2) = x_1 + x_2\ and\ g_2(x_1, x_2) = x_1 - x_2.$$

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

Joint Probability Distribution of Functions of Random Variables

Also, since the equations $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$ have $x_1 = (y_1 + y_2)/2, x_2 = (y_1 - y_2)/2$ as their solution, it follows from

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2)|J(x_1, x_2)|^{-1}, where J(x_1, x_2) = -2$$

Then

$$f_{Y_1 Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1, X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right)$$

For instance, if X_1 and X_2 are independent uniform (0,1) random variables, then

$$f_{Y_1 Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2} & 0 \le y_1 + y_2 \le 2, 0 \le y_1 - y_2 \le 2 \\ 0 & otherwise \end{cases}$$

Joint Probability Distribution of Functions of Random Variables

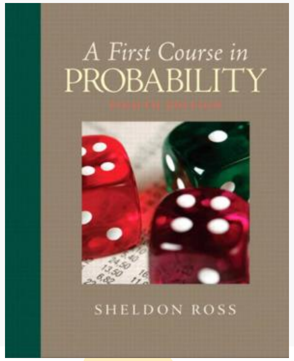
Or if X_1 and X_2 are independent **exponential random variables** with respective parameters λ_1 and λ_2, then

$$f_{Y_1Y_2}(y_1, y_2) = \begin{cases} \frac{\lambda_1\lambda_2}{2}\exp\{-\lambda_1\left(\frac{y_1+y_2}{2}\right) - \lambda_2\left(\frac{y_1-y_2}{2}\right)\} & y_1+y_2 \geq 0, y_1-y_2 \geq 0 \\ 0 & otherwise \end{cases}$$

Finally, if X_1 and X_2 are independent **standard normal random variables**, then

$$f_{Y_1Y_2}(y_1, y_2) = \frac{1}{4\pi}e^{-\left[\frac{(y_1+y_2)^2}{8}+\frac{(y_1-y_2)^2}{8}\right]} = \frac{1}{4\pi}e^{-\frac{(y_1^2+y_2^2)}{4}} = \frac{1}{\sqrt{4\pi}}e^{-\frac{y_1^2}{4}}\frac{1}{\sqrt{4\pi}}e^{-\frac{y_2^2}{4}}$$

[#12] Assignment



- Selected Problems from Sheldon Ross Textbook [1].

6.8. The joint probability density function of X and Y is given by

$$f(x, y) = c(y^2 - x^2)e^{-y} \qquad -y \le x \le y,\ 0 < y < \infty$$

(a) Find c.
(b) Find the marginal densities of X and Y.
(c) Find $E[X]$.

6.9. The joint probability density function of X and Y is given by

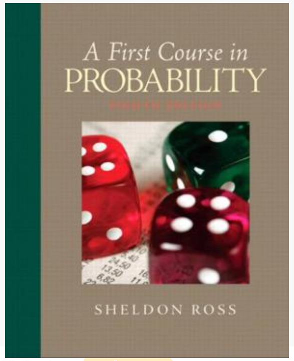
$$f(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) \qquad 0 < x < 1,\ 0 < y < 2$$

(a) Verify that this is indeed a joint density function.
(b) Compute the density function of X.
(c) Find $P\{X > Y\}$.
(d) Find $P\{Y > \frac{1}{2} | X < \frac{1}{2}\}$.
(e) Find $E[X]$.
(f) Find $E[Y]$.

6.12. The number of people that enter a drugstore in a given hour is a Poisson random variable with parameter $\lambda = 10$. Compute the conditional probability that at most 3 men entered the drugstore, given that 10 women entered in that hour. What assumptions have you made?

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

[#12] Assignment



6.40. The joint probability mass function of X and Y is given by

$$p(1,1) = \frac{1}{8} \qquad p(1,2) = \frac{1}{4}$$
$$p(2,1) = \frac{1}{8} \qquad p(2,2) = \frac{1}{2}$$

(a) Compute the conditional mass function of X given $Y = i, i = 1, 2$.

(b) Are X and Y independent?

(c) Compute $P\{XY \leq 3\}, P\{X + Y > 2\}, P\{X/Y > 1\}$.

6.41. The joint density function of X and Y is given by

$$f(x, y) = xe^{-x(y+1)} \quad x > 0, \ y > 0$$

(a) Find the conditional density of X, given $Y = y$, and that of Y, given $X = x$.

(b) Find the density function of $Z = XY$.

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

Reference

Ross, S. (2010). *A first course in probability*. Pearson.

The End

If you have any questions, please do not hesitate to ask me.

Thank you for your attention))