Probability 2 - Notes 9

The results for two random variables are now extended to n random variables.

Joint p.d.f. This is defined to be a function $f_{X_1,X_2,...,X_n}(x_1,...,x_n)$ such that for any measurable set A contained in $\Re^n$,

$$P((X_1,...,X_n)\in A)=\int...\int_{(x_1,...,x_n)\in A} f_{X_1,X_2,...,X_n}(x_1,...,x_n)dx_1...dx_n$$

It is convenient to write the joint p.d.f. using vector notation as $f_{\mathbf{X}}(\mathbf{x})$, where $\mathbf{X}$ and $\mathbf{x}$ are n-vectors with i^{th} entries X_i and x_i respectively.

Marginal p.d.f. Simply integrate out the variables which are not required.

Example $f_{X,Y,Z}(x,y,z)=3$ for $0<x<z$, $0<y<z$ and $0<z<1$. Then

$f_{X,Y}(x,y)=\int_{\max(x,y)}^{1}3dz=3(1-\max(x,y))$ for $0<x<1$ and $0<y<1$.

$f_{X,Z}(x,z)=\int_0^z 3dy=3z$ for $0<x<z<1$.

$f_{Y,Z}(y,z)=\int_0^z 3dx=3z$ for $0<y<z<1$.

Using the (marginal) joint p.d.f. for $f_{X,Z}(x,z)$, $f_X(x)=\int_x^1 3zdz=\frac{3}{2}(1-x^2)$ for $0<x<1$.

Using the (marginal) joint p.d.f. for $f_{Y,Z}(y,z)$, $f_Y(y)=\int_y^1 3zdz=\frac{3}{2}(1-y^2)$ for $0<y<1$.

Using the (marginal) joint p.d.f. for $f_{X,Z}(x,z)$, $f_Z(z)=\int_0^z 3zdx=3z^2$ for $0<z<1$.

Conditional p.d.f We can define the conditional p.d.f. for one set of random variables given another set, so for $1\leq m<n$,

$$f_{X_{m+1},...,X_n|X_1,...,X_m}(x_{m+1},\dots,x_n|x_1,\dots,x_m)=\frac{f_{X_1,...,X_n}(x_1,\dots,x_n)}{f_{X_1,...,X_m}(x_1,\dots,x_m)}$$

Example Consider the example above and condition on one random variable. We will consider two out of the three cases. For each $0<z<1$,

$$f_{X,Y|Z}(x,y,|z)=\frac{3}{3z^2}=\frac{1}{z^2}$$

for $0<x<z$, $0<y<z$. So $X,Y|Z=z$ are independent random variables each with $U(0,z)$ distribution. We say that they are conditionally independent.

For each $0<x<1$,

$$f_{Y,Z|X}(y,z|x) = \frac{3}{\frac{3}{2}(1-x^2)} = \frac{2}{(1-x^2)}$$

for $0 < y < z$ and $x < z < 1$.

Now consider conditioning on two random variables. Again we will consider two of the three cases. For each $0 < x < z < 1$,

$$f_{Y|X,Z}(y|x,z) = \frac{3}{3z} = \frac{1}{z}$$

for $0 < y < z$. So the conditional distribution of $Y|X = x, Z = z$ depends only on z and is $U(0,z)$.

Also for each $0 < x < 1$ and $0 < y < 1$,

$$f_{Z|X,Y}(z|x,y) = \frac{3}{3(1-\max(x,y))} = \frac{1}{(1-\max(x,y))}$$

for $\max(x,y) < z < 1$. Hence $Z|X = x, Y = y \sim U(\max(x,y),1)$.

Independence (Mutual) n jointly continuous random variables $X_1,...,X_n$ are said to be (mutually) independent if $f_{X_1,...X_n}(x_1,...,x_n) = \prod_{i=1}^n f_{X_i}(x_i)$ for all $x_1,...,x_n$.

Since the p.d.f. of any subset of the X_i is obtained by integrating out the other variables it immediately follows that

$$f_{X_{i_1},...X_{i_r}}(x_{i_1},...,x_{i_r}) = \prod_{j=1}^{r} f_{X_{i_j}}(x_{i_j})$$

for all $x_{1_1},...,x_{i_r}$ for all possible subsets $i_1,...,i_r$ and all $r = 2,...,n$.

Then for any events $'X_i \in A_i'$ $(i = 1,...,n)$ it is easily seen (by integrating over the appropriate sets) that

$$P('X_{i_j} \in A_{i_j}'\ j = 1,...,r) = \prod_{j=1}^{r} P('X_{i_j} \in A_{i_j}')$$

for all possible subsets of the n events, so that the events $'X_i \in A_i'$ $(i = 1,...,n)$ are mutually independent.

In addition if events $'X_i \in A_i'$ $(i = 1,...,n)$ are mutually independent for all such events, if we take $A_i = (x_i - dx_i, x_i]$ for $dx_i > 0$ small, then

$$f_{X_1,...X_n}(x_1,...,x_n)dx_1...dx_n \approxeq P('X_i \in A_i' \; i=1,...,n) = \prod_{i=1}^{n} P('X_i \in A_i') \approxeq \prod_{j=1}^{r} f_{X_{i_j}}(x_{i_j})dx_1...dx_n$$

It immediately follows that $X_1,...,X_n$ are independent.

Hence independence for the $X's$ is equivalent saying that all events $'X_i \in A_i'$, $(i=1,...,n)$, are mutually independent.

Properties.

1. If $X_1,...,X_n$ are independent then the joint p.d.f. is obtained by multiplying the individual p.d.f.'s together (for jointly continuous r.v.'s the joint p.d.f. is the 'likelihood' in statistics).

2. You can 'spot' independence in the same way as for two random variables. $X_1,...,X_n$ are independent iff the ranges are independent and $f_{X_1,...X_n}(x_1,...,x_n) = \prod_{i=1}^{n} g_i(x_i)$ for some function g_i. When this condition holds then the marginal p.d.f's are easily obtained. $f_{X_i}(x_i) = c_i g_i(x_i)$ for a suitable choice of $c_1,...,c_n$ with $\prod_{i=1}^{n} c_i = 1$.

3. If $X_1,...,X_n$ are independent then, for any functions h_i for which the expectations exist, $E\left[\prod_{i=1}^{n} h_i(X_i)\right] = \prod_{i=1}^{n} E[h_i(X_i)]$. This provides useful results for the m.g.f. (see examples).

Examples.

1. $f_{X,Y,Z}(x,y,z) = kxy^2 = (kx)(y^2)(1)$ for $0<x<1$, $0<y<1$, $0<z<1$. The ranges are independent and the joint p.d.f. splits as indicated. Hence X,Y,Z are independent and $f_X(x) = c_1 kx$ for $0<x<1$, $f_Y(y) = c_2 y^2$ for $0<y<1$ and $f_Z(z) = c_2$ for $0<z<1$, where $c_1c_2c_3 = 1$. We can find the constant k, c_1, c_2, c_3 from the results that each marginal p.d.f. integrates to 1 and $c_1c_2c_3 = 1$. Hence $c_3 = 1$, $c_2 = 3$, $kc_3 = 2$ and $c_1 = \frac{1}{3}$ and hence $k = 6$.

2. $X_1,...,X_n$ are independent with $X_j \sim Gamma(\theta, \alpha_j)$. Then we can use property 3 to show that $Y = \sum_{j=1}^{n} X_j \sim Gamma\left(\theta, \sum_{j=1}^{n} \alpha_j\right)$.

$$M_Y(t) = E\left[e^{t\sum_{j=1}^{n} X_j}\right] = E\left[\prod_{j=1}^{n} e^{tX_j}\right] = \prod_{j=1}^{n} M_{X_j}(t) = \prod_{j=1}^{n}\left(1-\frac{t}{\theta}\right)^{-\alpha_j} = \left(1-\frac{t}{\theta}\right)^{-\sum_{j=1}^{n}\alpha_j}$$

The result that $Y \sim Gamma\left(\theta, \sum_{j=1}^{n} \alpha_j\right)$ then follows from the uniqueness of the m.g.f.

3. $X_1,...,X_n$ are independent with $X_j \sim N(\mu_j, \sigma_j^2)$. Then we can use property 3 to show that $Y = \sum_{j=1}^{n} a_j X_j \sim N\left(\sum_{j=1}^{n} a_j\mu_j, \sum_{j=1}^{n} a_j^2\sigma_j^2\right)$.

$$\begin{aligned} M_Y(t) &= E\left[e^{t\sum_{j=1}^{n} a_j X_j}\right] = E\left[\prod_{j=1}^{n} e^{ta_jX_j}\right] = \prod_{j=1}^{n} M_{X_j}(a_jt) \\ &= \prod_{j=1}^{n} e^{\mu_j(a_jt)+(\sigma^2(a_jt)^2/2)} = e^{t\sum_{j=1}^{n} a_j\mu_j + (t^2/2)\sum_{j=1}^{n} a_j^2\sigma_j^2} \end{aligned}$$

The result that $Y \sim N\left(\sum_{j=1}^{n} a_j\mu_j, \sum_{j=1}^{n} a_j^2\sigma_j^2\right)$ then follows from the uniqueness of the m.g.f.

Transformations of variables

Let $X_1,...,X_n$ be n jointly continuous random variables with joint p.d.f. $f_{X_1,...,X_n}(x_1,...,x_n)$ which has support S contained in $\Re^n$. Consider random variables $Y_i = g_i(X_1,...,X_n)$ for $i = 1,...,n$ which is a one to one mapping from S to D with inverses $X_i = h_i(Y_1,...,Y_n)$ (for $i = 1,...,n$) which have continuous partial derivatives. Then

$$f_{Y_1,...,Y_n}(y_1,...,y_n) = f_{X_1,...,X_n}(h_1(y_1,...,y_n),...,h_n(y_1,...,y_n)) \times \left\| \begin{matrix} \frac{\partial h_1(y_1,...,y_n)}{\partial y_1} & \cdots & \frac{\partial h_1(y_1,...,y_n)}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n(y_1,...,y_n)}{\partial y_1} & \cdots & \frac{\partial h_n(y_1,...,y_n)}{\partial y_n} \end{matrix} \right\|$$

for $(y_1,...,y_n) \in D$. You can find D by rewriting the constraints on the ranges of $x_1,...,x_n$ in terms of $y_1,...,y_n$.

Example. X_1,X_2,X_3 are independent $Exp(\theta)$. So $f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \theta^3 e^{-\theta(x_1+x_2+x_3)}$ for $x_1 > 0$, $x_2 > 0$ and $x_3 > 0$. Find the joint p.d.f. for $Y_1 = X_1$, $Y_2 = X_1 + X_2$ and $Y_3 = X_1 + X_2 + X_3$.

The inverses are $X_1 = Y_1$, $X_2 = Y_2 - Y_1$ and $X_3 = Y_3 - Y_2$. Hence using the result above:

$$f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = \theta^3 e^{-\theta y_3} \times \left\| \begin{matrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{matrix} \right\| = \theta^3 e^{-\theta y_3}$$

The ranges $x_1 > 0$, $x_2 > 0$ and $x_3 > 0$ become $y_1 > 0$, $y_2 - y_1 > 0$ and $y_3 - y_2 > 0$, i.e. $0 < y_1 < y_2 < y_3 < \infty$.

The joint moment generating function.

The joint m.g.f. for n random variables $X_1,...,X_n$ is now defined and its properties given. Let $\mathbf{X}$ and $\mathbf{t}$ be n-vectors (column vectors) with j^{th} entries X_j and t_j respectively. Then

$$M_{\mathbf{X}}(\mathbf{t}) = M_{X_1,...,X_n}(t_1,...,t_n) = E\left[e^{\sum_{j=1}^{n} t_j X_j}\right] = E[e^{\mathbf{t}^T\mathbf{X}}]$$

Properties.

1. The joint m.g.f. of a subset $X_{i_1},...X_{i_r}$ of the $X's$ is obtained by setting $t_j = 0$ for all j not in the set $\{i_1,...,i_r\}$. Note that the joint m.g.f. equals one when $t_j = 0$ for all $j = 1,...,n$ (i.e. $M_{\mathbf{X}}(\mathbf{0}) = 1$).

2. If $X_1,...,X_n$ are independent then

$$M_{\mathbf{X}}(\mathbf{t}) = E\left[e^{\sum_{j=1}^{n} t_j X_j}\right] = E\left[\prod_{j=1}^{n} e^{t_j X_j}\right] = \prod_{j=1}^{n} M_{X_j}(t_j)$$

3. There is a unique relationship between the joint p.d.f. and the joint m.g.f. (so one determines the other).

4. If $M_{\mathbf{X}}(\mathbf{t}) = \prod_{j=1}^{n} g_j(t_j)$ for some functions g_j, $j = 1,...,n$, then $X_1,...,X_n$ are independent.

Proof. If we set $t_i = 0$ for $i \neq j$, then we obtain the m.g.f. for X_j, hence $M_{X_j}(t_j) = g_j(t_j)\prod_{i\neq j} g_i(0)$. Also setting $t_i = 0$ for $i = 1,...,n$ gives $1 = \prod_{i=1}^{n} g_i(0)$. Therefore $M_{X_j}(t_j) = \frac{g_j(t_j)}{g_j(0)}$ and hence

$$M_{\mathbf{X}}(\mathbf{t}) = \prod_{j=1}^{n} g_j(t_j) = \prod_{j=1}^{n} g_j(0) M_{X_j}(t_j) = \prod_{j=1}^{n} M_{X_j}(t_j)$$

Hence from property 3 of the joint m.g.f., $X_1,...,X_n$ are independent with the p.d.f. of X_j determined by the m.g.f. $M_{X_j}(t_j) = \frac{g_j(t_j)}{g_j(0)}$.

Use of the joint m.g.f. to obtain some important results in statistics.

1. If $X_1,...,X_n$ are independent $N(\mu, \sigma^2)$ and if $Z_j = \frac{X_j - \mu}{\sigma}$ for $j = 1,...,n$, then $Z_1,...,Z_n$ are independent $N(0,1)$.

Proof.

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) &= E\left[e^{\sum_{j=1}^{n} t_j (X_j - \mu)/\sigma}\right] = \prod_{j=1}^{n} E[e^{t_j (X_j - \mu)/\sigma}] = \prod_{j=1}^{n}\left(e^{-\mu t_j/\sigma} M_{X_j}(t_j/\sigma)\right) \\ &= \prod_{j=1}^{n}\left(e^{-\mu t_j/\sigma} e^{\mu(t_j/\sigma) + (\sigma^2/2)(t_j \sigma)^2}\right) = \prod_{j=1}^{n} e^{t_j^2/2} \end{aligned}$$

Hence by property 4 of the joint m.g.f., $Z_1,...,Z_n$ are independent with $M_{Z_j}(t_j) = e^{t_j^2/2}$, which is the m.g.f. of the $N(0,1)$ distribution. Hence from the uniqueness property of the m.g.f., $Z_1,...,Z_n$ are independent $N(0,1)$.

2. If $Z_1,...,Z_n$ are independent $N(0,1)$ and $\mathbf{Y} = \mathbf{A}\mathbf{Z}$ with $\mathbf{Z}$ the n-vector with j^{th} entry Z_j and $\mathbf{A}$ an $n \times n$ orthogonal matrix (i.e. $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$ where $\mathbf{I}$ is the $n \times n$ identity matrix), then $\mathbf{Y}$ is an n-vector (entries $Y_1,...,Y_n$) of independent $N(0,1)$ random variables.

Proof. Now $M_{\mathbf{Z}}(\mathbf{t}) = \prod_{j=1}^{n} M_{Z_j}(t_j) = \prod_{j=1}^{n} e^{t_j^2/2} = e^{(1/2)\mathbf{t}^T\mathbf{t}}$. Hence

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E[e^{\mathbf{t}^T\mathbf{Y}}] = E\left[e^{\mathbf{t}^T\mathbf{A}\mathbf{Z}}\right] = E\left[e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{Z}}\right] = M_{\mathbf{Z}}(\mathbf{A}^T\mathbf{t}) \\ &= e^{(1/2)(\mathbf{A}^T\mathbf{t})^T(\mathbf{A}^T\mathbf{t})} = e^{(1/2)\mathbf{t}^T\mathbf{A}\mathbf{A}^T\mathbf{t}} = e^{(1/2)\mathbf{t}^T\mathbf{t}} = \prod_{j=1}^{n} e^{t_j^2/2} \end{aligned}$$

Hence using property 4 of the joint m.g.f. and the uniqueness of the m.g.f., $Y_1,...,Y_n$ are independent $N(0,1)$.

3. If $Y_1,...,Y_n$ are independent $N(0,1)$ then Y_1 and $U = \sum_{j=2}^n Y_j^2$ are independent with $Y_1 \sim N(0,1)$ and $U \sim \chi^2_{(n-1)}$.

Proof. We use the result proved earlier that if $Y \sim N(0,1)$ then $E[e^{tY^2}] = (1-2t)^{-1/2}$. Now

$$\begin{aligned} M_{Y_1,U}(s,t) &= E\left[e^{sY_1+t\sum_{j=2}^n Y_j^2}\right] = E\left[e^{sY_1}\prod_{j=2}^n e^{tY_j^2}\right] = E[e^{sY_1}]\prod_{j=2}^n E[e^{tY_j^2}] \\ &= M_{Y_1}(s)\prod_{j=2}^n E[e^{tY_j^2}] = e^{s^2/2}\prod_{j=2}^n (1-2t)^{-1/2} = e^{s^2/2}(1-2t)^{-(n-1)/2} \end{aligned}$$

Hence using property 4 of the joint m.g.f and the uniqueness of the m.g.f., $Y_1 \sim N(0,1)$ independent of $U \sim \chi^2_{(n-1)}$.

Theorem. If $X_1,...,X_n$ are independent $N(\mu,\sigma^2)$ then $\sqrt{n}(\overline{X}-\mu)/\sigma \sim N(0,1)$ independent of $\sum_{j=1}^n (X_j-\overline{X})^2/\sigma^2 \sim \chi^2_{(n-1)}$.

Proof. Let $Z_j = (X_j-\mu)/\sigma$ for $j = 1,...,n$. Then $\sqrt{n}\overline{Z} = \sqrt{n}(\overline{X}-\mu)/\sigma$ and $\sum_{j=1}^n (Z_j-\overline{Z})^2 = \sum_{j=1}^n (X_j-\overline{X})^2/\sigma^2$. Also from result (1) $Z_1,...,Z_n$ are independent $N(0,1)$.

Use result (2) with $\mathbf{A}$ the $n\times n$ matrix with first row $\left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{n}},...,\frac{1}{\sqrt{n}}\right)$. Then $Y_1 = \left(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{n}},...,\frac{1}{\sqrt{n}}\right)\mathbf{Z} = \sqrt{n}\overline{Z}$. Also

$$\sum_{j=1}^n Y_j^2 = \mathbf{Y}^T\mathbf{Y} = (\mathbf{AZ})^T(\mathbf{AZ}) = \mathbf{Z}^T\mathbf{A}^T\mathbf{AZ} = \mathbf{Z}^T\mathbf{Z} = \sum_{j=1}^n Z_j^2$$

Therefore

$$\sum_{j=1}^n (Z_j-\overline{Z})^2 = \sum_{j=1}^n Z_j^2 - n\overline{Z}^2 = \sum_{j=1}^n Y_j^2 - Y_1^2 = \sum_{j=2}^n Y_j^2$$

Then from result (2) $Y_1,...,Y_n$ are independent $N(0,1)$ and from result (3) $Y_1 = \sqrt{n}\overline{Z} = \sqrt{n}(\overline{X}-\mu)/\sigma \sim N(0,1)$ independent of $U = \sum_{j=2}^n Y_j^2 = \sum_{j=1}^n (Z_j-\overline{Z})^2 = \sum_{j=1}^n (X_j-\overline{X})^2/\sigma^2 \sim \chi^2_{(n-1)}$.

Note: This provides the basis for the t and χ^2 tests met in Fundamentals of Statistics 1. Orthogonal transformations of independent $N(0,1)$ variables will also be used to prove results in Statistical Modelling 1.