2.7. MULTIVARIATE EXPECTATIONS AND COVARIANCE

Consider vector random variable $\mathbf{X} = (X_1, ..., X_k)$ with range $\mathbb{X}^{(k)}$ and joint mass or density function $f_{X_1,...,X_k}$, and a real-valued scalar function g whose domain includes $\mathbb{X}^{(k)}$.

DEFINITION

For random variables $X_1, ..., X_k$ with range $\mathbb{X}^{(k)}$ with mass/density function $f_{X_1,...,X_k}$, the **expectation** of $g(X_1, ..., X_k)$ is defined in the discrete and continuous cases by

$$\mathrm{E}_{f_{X_1,...,X_k}}[\ g(X_1, ..., X_k)\ ] = \begin{cases} \displaystyle\sum_{x_1} ... \sum_{x_k} g(x_1, ..., x_k) f_{X_1,...,X_k}(x_1, ..., x_k) \\ \displaystyle\int_{\mathbb{X}_1} ... \int_{\mathbb{X}_k} g(x_1, ..., x_k) f_{X_1,...,X_k}(x_1, ..., x_k)\ dx_1 ... dx_k \end{cases}$$

Properties of Multivariate Expectations :

(i) Let $X_1, ...X_k$ be random variables with joint mass function/pdf $f_{X_1,...,X_k}$. Let g and h be real-valued functions whose domains include $\mathbb{X}^{(k)}$, and let a and b be constants. Then, if $f_{\mathbf{X}} \equiv f_{X_1,...,X_k}$,

$$\mathrm{E}_{f_{\mathbf{X}}}[\ ag(X_1, ..., X_k) + bh(X_1, ..., X_k)\ ] = a\mathrm{E}_{f_{\mathbf{X}}}[\ g(X_1, ..., X_k)\ ] + b\mathrm{E}_{f_{\mathbf{X}}}[\ h(X_1, ..., X_k)\ ].$$

(ii) Let $X_1, ...X_k$ be independent random variables with mass functions/pdfs $f_{X_1}, ..., f_{X_k}$ respectively. Let $g_1, ..., g_k$ be scalar functions of $X_1, ..., X_k$ respectively (that is, g_i is a function of X_i *only* for $i = 1, ..., k$). If $g(X_1, ..., X_k) = g_1(X_1)...g_k(X_k)$, then

$$\mathrm{E}_{f_{\mathbf{X}}}[\ g(X_1, ..., X_k)\ ] \ = \ \prod_{i=1}^{k} \mathrm{E}_{f_{X_i}}[\ g_i(X_i)\ ]$$

where $\mathrm{E}_{f_{X_i}}[\ g_i(X_i)\ ]$ is the marginal expectation of $g_i(X_i)$ with respect to f_{X_i}.

COVARIANCE AND CORRELATION

DEFINITION

If X_1 and X_2 are random variables with joint mass function/pdf f_{X_1,X_2}, then the **covariance** of X_1 and X_2 is denoted $\mathrm{Cov}_{f_{X_1,X_2}}[X_1, X_2]$, and is defined by

$$\mathrm{Cov}_{f_{X_1,X_2}}[X_1, X_2] = \mathrm{E}_{f_{X_1,X_2}}[(X_1 - \mu_1)(X_2 - \mu_2)] = \mathrm{E}_{f_{X_1,X_2}}[X_1 X_2] - \mu_1 \mu_2$$

where $\mu_i = \mathrm{E}_{f_{X_i}}[X_i]$ is the marginal expectation of X_i, for $i = 1, 2$, and where $\mathrm{E}_{f_{X_1,X_2}}[X_1 X_2]$ is the expectation of function $g(x_1, x_2) = x_1 x_2$ with respect to the joint distribution f_{X_1,X_2}.

DEFINITION

If X_1 and X_2 are random variables with joint mass function/pdf f_{X_1,X_2}, then the **correlation** of X_1 and X_2 is denoted $\mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2]$, and is defined by

$$\mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2] = \frac{\mathrm{Cov}_{f_{X_1,X_2}}[X_1, X_2]}{\sqrt{\mathrm{Var}_{f_{X_1}}[X_1]\ \mathrm{Var}_{f_{X_2}}[X_2]}}$$

Note: if $\mathrm{Cov}_{f_{X_1,X_2}}[X_1, X_2] = \mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2] = 0$. then variables X_1 and X_2 are *uncorrelated.*

THEOREM

If random variables X_1 and X_2 are *independent* then

$$\mathrm{Cov}_{f_{X_1,X_2}}[X_1, X_2] = 0$$

and X_1 and X_2 are uncorrelated.

PROOF

If X_1 and X_2 are independent, then $\mathrm{E}_{f_{X_1,X_2}}[X_1X_2] = \mathrm{E}_{f_{X_1}}[X_1]\mathrm{E}_{f_{X_2}}[X_2]$, and hence the result follows.

NOTES:

(i) For random variables X_1 and X_2, with (marginal) expectations μ_1 and μ_2 respectively, and (marginal) variances σ_1^2 and σ_2^2 respectively, if random variables Z_1 and Z_2 are defined

$$Z_1 = (X_1 - \mu_1)/\sigma_1 \qquad Z_2 = (X_2 - \mu_2)/\sigma_2$$

that is, Z_1 and Z_2 are *standardized* variables. Then $\mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2] = \mathrm{Cov}_{f_{Z_1,Z_2}}[Z_1, Z_2]$.

(ii) Extension to k variables: covariances can only be calculated for *pairs* of random variables, but if k variables have a joint probability structure it is possible to construct a $k \times k$ *matrix*, $\mathbf{C}$ say, of covariance values, whose (i, j)th element is

$$\mathrm{Cov}_{f_{X_i,X_j}}[X_i, X_j]$$

for $i, j = 1, .., k$, that captures the complete covariance structure in the joint distribution. If $i \neq j$, then

$$\mathrm{Cov}_{f_{X_j,X_i}}[X_j, X_i] = \mathrm{Cov}_{f_{X_i,X_j}}[X_i, X_j]$$

so $\mathbf{C}$ is *symmetric*, and if $i = j$,

$$\mathrm{Cov}_{f_{X_i,X_i}}[X_i, X_i] \equiv \mathrm{Var}_{f_{X_i}}[X_i]$$

The matrix $\mathbf{C}$ is referred to as the **variance-covariance** matrix.

(iii) If random variable X is defined by

$$X = \sum_{i=1}^{k} a_i X_i$$

for random variables $X_1, ..., X_k$ and constants $a_1, ..., a_k$, then

$$\mathrm{E}_{f_X}[X] = \sum_{i=1}^{k} a_i \mathrm{E}_{f_{X_i}}[X_i] \qquad \mathrm{Var}_{f_X}[X] = \sum_{i=1}^{k} a_i^2 \mathrm{Var}_{f_{X_i}}[X_i] + 2\sum_{i=1}^{k}\sum_{j=1}^{i-1} a_i a_j \mathrm{Cov}_{f_{X_i,X_j}}[X_i, X_j]$$

(iv) Combining (i) and (iii) when $k = 2$, and defining standardized variables Z_1 and Z_2,

$$\begin{aligned} 0 \leq \mathrm{Var}_{f_{Z_1,Z_2}}[Z_1 \pm Z_2] &= \mathrm{Var}_{f_{Z_1}}[Z_1] + \mathrm{Var}_{f_{Z_2}}[Z_2] \pm 2\ \mathrm{Cov}_{f_{Z_1,Z_2}}[Z_1, Z_2] \\ &= 1 + 1 \pm 2\ \mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2] \\ &= 2(1 \pm \mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2]) \end{aligned}$$

and hence $-1 \leq \mathrm{Corr}_{f_{X_1,X_2}}[X_1, X_2] \leq 1$.

2.8. SPECIAL FUNCTIONS I - SUMS OF RANDOM VARIABLES

THEOREM - THE CONVOLUTION FORMULA

Let X_1 and X_2 be continuous random variables taking values on $\mathbb{X}_1$ and $\mathbb{X}_2$ respectively with joint density function f_{X_1,X_2}. Then, if random variable Y is defined by $Y = X_1 + X_2$, the marginal density of Y is given by

$$f_Y(y) = \int f_{X_1,X_2}(x_1, y - x_1)dx_1$$

PROOF

Consider the multivariate transformation $\mathbf{g} : (X_1, X_2) \longrightarrow (Y, Z)$, where $(Y, Z) = (X_1 + X_2, X_1)$, so that the inverse transformation is given by $(X_1, X_2) = (Z, Y - Z)$. Then by the previous theorem,

$$f_{Y,Z}(y, z) = f_{X_1,X_2}(z, y - z)J(y, z)$$

for

$$(y, z) \in B = \{(y, z)|(y, z) = \mathbf{g}(x_1, x_2), x_1 \in \mathbb{X}_1, x_2 \in \mathbb{X}_2\}.$$

Recall that $J(y, z)$ is a 2×2 matrix of partial derivatives

$$\begin{bmatrix} \dfrac{\partial}{\partial t_1}\left\{{g_1}^{-1}(t_1, t_2)\right\}_{t_1=y,t_2=z} & \dfrac{\partial}{\partial t_2}\left\{{g_1}^{-1}(t_1, t_2)\right\}_{t_1=y,t_2=z} \\ \dfrac{\partial}{\partial t_1}\left\{{g_2}^{-1}(t_1, t_2)\right\}_{t_1=y,t_2=z} & \dfrac{\partial}{\partial t_2}\left\{{g_2}^{-1}(t_1, t_2)\right\}_{t_1=y,t_2=z} \end{bmatrix}$$

Here, ${g_1}^{-1}(t_1, t_2) = t_2$, ${g_1}^{-1}(t_1, t_2) = t_1 - t_2$. Hence

$$J(y, z) = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = |-1|$$

so $J(y, z) = 1$. Therefore

$$f_{Y,Z}(y, z) = f_{X_1,X_2}(z, y - z) \qquad (y, z) \in B$$

so

$$f_Y(y) = \int_{\mathbb{Z}} f_{Y,Z}(y, z)\ dz = \int_{\mathbb{X}_1} f_{X_1,X_2}(x_1, y - x_1)\ dx_1$$

as $X_1 \equiv Z$.

Special Case : If X_1 and X_2 are **independent**, then

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

so if $Y = X_1 + X_2$, then Y has pdf given by

$$f_Y(y) = \int_{\mathbb{X}_1} f_{X_1}(x_1)f_{X_2}(y - x_1)\ dx_1$$

2.9. SPECIAL FUNCTIONS II - ORDER STATISTICS

DEFINITION

For k random variables $X_1, ..., X_k$, the **order statistics**, $Y_1, ..., Y_k$, are defined by

$$Y_i = X_{(i)} \text{ - "the } i\text{th smallest value in } X_1, ..., X_k\text{"}$$

for $i = 1, ..., k$, so that $Y_1 = X_{(1)} = \text{Min}\{X_1, ..., X_k\}$, $Y_k = X_{(k)} = \text{Max}\{X_1, ..., X_k\}$.
For k independent, identically distributed random variables $X_1, ..., X_k$, with marginal density function f_X, there are three results to consider;

RESULT 1 The joint density function of the order statistics $Y_1, ..., Y_k$ is given by

$$f_{Y_1,...,Y_k}(y_1, ..., y_k) = k! f_X(y_1)...f_X(y_k) \qquad y_1 < ... < y_k$$

RESULT 2 The marginal pdf of the jth order statistic Y_j for $j = 1, ..., k$ has the form

$$f_{Y_j}(y_j) = \frac{k!}{(j-1)!(k-j)!} \{F_X(y_j)\}^{j-1} \{1 - F_X(y_j)\}^{k-j} f_X(y_j)$$

Special Case: To derive the marginal pdf of Y_k, first consider the marginal cdf of Y_k;

$$\begin{aligned} F_{Y_k}(y_k) &= \mathrm{P}[Y_k \le y_k] = \mathrm{P}[\max\{X_1, ..., X_k\} \le y_k] = \mathrm{P}[X_1 \le y_k, X_2 \le y_k, ..., X_k \le y_k] \\ &= \prod_{i=1}^{k} \mathrm{P}[X_i \le y_k] = \prod_{i=1}^{k} \{F_X(y_k)\} = \{F_X(y_k)\}^k \end{aligned}$$

Now, differentiate $F_{Y_k}(y_k)$ with respect to y_k to get the marginal pdf of Y_k,

$$f_{Y_k}(y_k) = k\{F_X(y_k)\}^{k-1} f_X(y_k)$$

RESULT 3 The joint pdf of the ith and jth $(i < j)$ order statistics Y_i and Y_j has the form

$$f_{Y_i,Y_j}(y_i, y_j) = \frac{k! f_X(y_i) f_X(y_j)}{(i-1)!(j-i-1)!(k-j)!} \{F_X(y_i)\}^{i-1} [F_X(y_j) - F_X(y_i)]^{j-i-1} \{1 - F_X(y_j)\}^{k-j}$$

for $y_i < y_j$.

Special Case: Setting $i = 1$ and $j = k$ in Result 3;

$$f_{Y_1,Y_k}(y_1, y_k) = \frac{k!}{(k-2)!} f_X(y_1) [F_X(y_k) - F_X(y_1)]^{k-2} f_X(y_k)$$

Using this joint pdf, the marginal pdf of the random variable $R = Y_k - Y_1$ (the **range**) can be obtained by transformation of variables techniques.

Result 1 can be proved by considering the transformation from $(X_1, ..., X_k)$ to $(Y_1, ..., Y_k)$; the result is easy to prove in the discrete case. Results 2 and 3 follow by marginalizing the joint mass/density derived in Result 1.

2.10. GENERATING FUNCTIONS

DEFINITION

For random variable X with mass/density function f_X, the **moment generating function**, or **mgf**, of X, M_X, is defined by

$$M_X(t) = \mathrm{E}_{f_X}[e^{tX}]$$

if this expectation exists for all values of $t \in (-h, h)$ for some $h > 0$, that is,

Discrete case $M_X(t) = \displaystyle\sum_{x \in \mathbb{X}} e^{tx} f_X(x)$ **Continuous case** $M_X(t) = \displaystyle\int_{x \in \mathbb{X}} e^{tx} f_X(x)\, dx$

It can be shown that if X_1 and X_2 are random variables taking values on $\mathbb{X}$ with mass/density functions f_{X_1} and f_{X_2}, and mgfs M_{X_1} and M_{X_2} respectively, then

$$f_{X_1}(x) \equiv f_{X_2}(x),\ x \in \mathbb{X} \quad \Longleftrightarrow \quad M_{X_1}(t) \equiv M_{X_2}(t),\ t \in (-h, h)$$

PROPERTIES OF MGFS

(i) If X is a **discrete** random variable, the rth derivative of M_X evaluated at t, $M_X^{(r)}(t)$, is given by

$$M_X^{(r)}(t) = \frac{d^r}{ds^r}\left\{M_X(s)\right\}_{s=t} = \frac{d^r}{ds^r}\left\{\sum_{x \in \mathbb{X}} e^{sx} f_X(x)\right\}_{s=t} = \sum_{x \in \mathbb{X}} x^r e^{tx} f_X(x)$$

and hence

$$M_X^{(r)}(0) = \sum_{x \in \mathbb{X}} x^r f_X(x) = \mathrm{E}_{f_X}[X^r]$$

If X is a **continuous** random variable, the rth derivative of M_X is given by

$$M_X^{(r)}(t) = \frac{d^r}{ds^r}\left\{\int_{x \in \mathbb{X}} e^{sx} f_X(x)\ d_x\right\}_{s=t} = \int_{x \in \mathbb{X}} x^r e^{tx} f_X(x)\ dx$$

and hence

$$M_X^{(r)}(0) = \int_{x \in \mathbb{X}} x^r f_X(x)\ dx = \mathrm{E}_{f_X}[X^r]$$

(ii) If X is a **discrete** random variable, then

$$M_X(t) = \sum_{x \in \mathbb{X}} e^{tx} f_X(x) = \sum_{x \in \mathbb{X}} \left\{\sum_{r=0}^{\infty} \frac{(tx)^r}{r!}\right\} f_X(x) = 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!}\left\{\sum_{x \in \mathbb{X}} x^r f_X(x)\right\} = 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} \mathrm{E}_{f_X}[X^r]$$

(similar result holds for the **continuous** case).

(iii) From the general result for expectations of functions of random variables

$$\mathrm{E}_{f_Y}[e^{tY}] \equiv \mathrm{E}_{f_X}[e^{t(aX+b)}] \Longrightarrow M_Y(t) = \mathrm{E}_{f_X}[e^{t(aX+b)}] = e^{bt}\mathrm{E}_{f_X}[e^{atX}] = e^{bt} M_X(at).$$

Therefore, if $Y = aX + b$, $M_Y(t) = e^{bt} M_X(at)$

THEOREM

Let $X_1, ..., X_k$ be independent random variables with mgfs $M_{X_1}, ..., M_{X_k}$ respectively. Then if random variable Y is defined by $Y = X_1 + ... + X_k$,

$$M_Y(t) = \prod_{i=1}^{k} M_{X_i}(t)$$

Proof Using the general result for expectations of functions of independent random variables,

$$M_Y(t) = \mathrm{E}_{f_Y}[e^{tY}] = \mathrm{E}_{f_{X_1,...,X_k}}[e^{t(X_1+...+X_k)}] = \prod_{i=1}^{k} \mathrm{E}_{f_{X_i}}[e^{tX_i}] = \prod_{i=1}^{k} M_{X_i}(t).$$

Special Case If $X_1, ..., X_k$ are **identically distributed**, then $M_{X_i}(t) \equiv M_X(t)$, say, for all i, so

$$M_Y(t) = \prod_{i=1}^{k} M_X(t) = \{M_X(t)\}^k$$

DEFINITION

For random variable X, with mass/density function f_X, the **factorial moment** or **probability generating function**, **fmgf** or **pgf** , of X, G_X, is defined by

$$G_X(t) = \mathrm{E}_{f_X}[t^X] = \mathrm{E}_{f_X}[e^{X \log t}] = M_X(\log t)$$

if this expectation exists for all values of $t \in (1-h, 1+h)$ for some $h > 0$.

Properties :

(i) Using similar techniques to those used for the mgf, it can be shown that

$$G_X^{(r)}(t) = \frac{d^r}{ds^r}\{G_X(s)\}_{s=t} = \mathrm{E}_{f_X}[X(X-1)...(X-r+1)t^{X-r}] \Longrightarrow G_X^{(r)}(1) = \mathrm{E}_{f_X}[X(X-1)...(X-r+1)]$$

where $\mathrm{E}_{f_X}[X(X-1)...(X-r+1)]$ is the rth **factorial moment**.

(ii) For **discrete** random variables, it can be shown by using a Taylor series expansion of G_X that, for $r = 1, 2, ...,$

$$\frac{G_X^{(r)}(0)}{r!} = \mathrm{P}[X = r]$$

DEFINITION

For random variable X with mass/density function f_X, the **cumulant generating function** of X, K_X, is defined by

$$K_X(t) = \log[M_X(t)]$$

for $t \in (-h, h)$ for some $h > 0$. The **characteristic function**, or **cf**, of X, C_X, is defined by

$$C_X(t) = \mathrm{E}_{f_X}[e^{itX}]$$

if this expectation exists for $t \in \mathbb{R}$.