EXPECTATIONS

DEFINITION

For a random variable X with range $\mathbb{X}$ with probability mass/density function f_X, the **expectation** or **expected value** of X with respect to f_X is defined by

$$\mathrm{E}_{f_X}[X] = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} x f_X(x) \quad = \quad \sum_{x \in \mathbb{X}} x f_X(x) & \text{DISCRETE} \\[2ex] \displaystyle\int_{-\infty}^{\infty} x f_X(x)\ dx \quad = \quad \int_{\mathbb{X}} x f_X(x)\ dx & \text{CONTINUOUS} \end{cases}$$

NOTE:
The expected value is a weighted average of the values of X, but the sum/integral may not be convergent, and hence the expected value may not be finite.

Extension
If g is a real-valued function whose domain includes $\mathbb{X}$, then

$$\mathrm{E}_{f_X}[g(X)] = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} g(x) f_X(x) \quad = \quad \sum_{x \in \mathbb{X}} g(x) f_X(x) & \text{DISCRETE} \\[2ex] \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x)\ dx \quad = \quad \int_{\mathbb{X}} g(x) f_X(x)\ dx & \text{CONTINUOUS} \end{cases}$$

PROPERTIES OF EXPECTATIONS

Let X be a random variable with mass function/pdf f_X. Let g and h be real-valued functions whose domains include $\mathbb{X}$, and let a and b be constants. Then

$$\mathrm{E}_{f_X}[ag(X) + bh(X)] = a\mathrm{E}_{f_X}[g(X)] + b\mathrm{E}_{f_X}[h(X)]$$

as (in the continuous case)

$$\begin{aligned} \mathrm{E}_{f_X}[ag(X) + bh(X)] &= \int [ag(x) + bh(x)] f_X(x)\ dx \ = \ a\int g(x) f_X(x)\ dx + b\int h(x) f_X(x)\ dx \\ &= a\mathrm{E}_{f_X}[g(X)] + b\mathrm{E}_{f_X}[h(X)] \end{aligned}$$

Special Cases :

(i) For constants a and b

$$\mathrm{E}_{f_X}[aX + b] = a\mathrm{E}_{f_X}[X] + b$$

(ii) Consider $g(x) = (x - \mathrm{E}_{f_X}[X])^2$. Write $\mu = \mathrm{E}_{f_X}[X]$ (a constant that does not depend on x). Then

$$\begin{aligned}
\mathrm{E}_{f_X}[g(X)] &= \int (x-\mu)^2 f_X(x)\ dx = \int x^2 f_X(x)\ dx\ - 2\mu \int x f_X(x)\ dx + \mu^2 \int f_X(x)\ dx \\
&= \int x^2 f_X(x)\ dx - 2\mu^2 + \mu^2\ = \int x^2 f_X(x)\ dx - \mu^2\ = \mathrm{E}_{f_X}[X^2] - \{\mathrm{E}_{f_X}[X]\}^2
\end{aligned}$$

We typically write

$$\mathrm{Var}_{f_X}[X] = \mathrm{E}_{f_X}[X^2] - \{\mathrm{E}_{f_X}[X]\}^2$$

and say that $\mathrm{Var}_{f_X}[X]$ is the **variance** of the distribution. Note also that

$$\mathrm{E}_{f_X}[X^2] = \mathrm{E}_{f_X}[X(X-1)] + \mathrm{E}_{f_X}[X]$$

so that

$$\mathrm{Var}_{f_X}[X] = \mathrm{E}_{f_X}[X(X-1)] + \mathrm{E}_{f_X}[X] - \{\mathrm{E}_{f_X}[X]\}^2$$

Typically we also consider $\sqrt{\mathrm{Var}_{f_X}[X]}$, which is the **standard deviation** of the distribution.

(iii) Consider $g(x) = x^k$ for $k = 1, 2, ...$ Then in the continous case

$$\mathrm{E}_{f_X}[g(X)] = \mathrm{E}_{f_X}[X^k] = \int x^k f_X(x)\ dx,$$

and $\mathrm{E}_{f_X}[X^k]$ is the kth **moment** of the distribution.

(iv) Consider $g(x) = (x-\mu)^k$ for $k = 1, 2, ...$ Then

$$\mathrm{E}_{f_X}[g(X)] = \mathrm{E}_{f_X}[(X-\mu)^k] = \int (x-\mu)^k f_X(x)\ dx,$$

and $\mathrm{E}_{f_X}[(X-\mu)^k]$ is the kth **central moment** of the distribution.

(v) Consider $g(x) = aX + b$. Then

$$\mathrm{Var}_{f_X}[aX+b] = a^2 \mathrm{Var}_{f_X}[X]$$

as

$$\begin{aligned}
\mathrm{Var}_{f_X}[g(X)] &= \mathrm{E}_{f_X}[(aX + b - \mathrm{E}_{f_X}[aX+b])^2] = \mathrm{E}_{f_X}[(aX + b - a\mathrm{E}_{f_X}[X] - b)^2] \\
&= \mathrm{E}_{f_X}[(a^2(X - \mathrm{E}_{f_X}[X])^2] = a^2 \mathrm{Var}_{f_X}[X]
\end{aligned}$$

The expectation is a measure of **location** of the distribution

The variance is a measure of **spread** or **scale** of the distribution

PROBABILITY AND MOMENT GENERATING FUNCTIONS

We have seen previously the definition of the **probability generating function**, or pgf, for a discrete random variable X as

$$G_X(t) = \sum_{x \in \mathbb{X}} t^x f_X(x).$$

We can now see that the pgf is actually the expected value of function $g(x) = t^x$ for real constant t, that is, we have

$$G_X(t) = \mathrm{E}_{f_X}\left[\ t^X\ \right].$$

The pgf can be used to identify a distribution if we do not have access to the mass function, and to find the distribution of the sum of independent discrete random variables.

The pgf is only defined for discrete random variables. We now study an analogous function that may be defined for both discrete and continuous random variables.

<u>DEFINITION</u>

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,

$$\textbf{Discrete case} \qquad M_X(t) = \sum_{x \in \mathbb{X}} e^{tx} f_X(x)$$

$$\textbf{Continuous case} \qquad M_X(t) = \int_{x \in \mathbb{X}} e^{tx} f_X(x)\ dx$$

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

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

that is, a probability distribution is uniquely defined by its mgf.

Note also that

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

so that G_X and M_X are closely related.

We saw examples of pgf calculations previously, and it is straightforward to establish that the pgf and the mgf can be routinely calculated for most distributions.

PROPERTIES OF PGFS AND MGFS

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

$$M_X^{(r)}(t) = \frac{d^r}{ds^r}\left\{M_X(s)\right\}_{s=t} = \sum_{x\in\mathbb{X}} x^r e^{tx} f_X(x) \Longrightarrow M_X^{(r)}(0) = \sum_{x\in\mathbb{X}} x^r f_X(x) = \mathrm{E}_{f_X}[X^r]$$

A similar result holds for the pgf, G_X; we have

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

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

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

$$M_X^{(r)}(t) = \frac{d^r}{ds^r}\left\{\int_{x\in\mathbb{X}} e^{sx} f_X(x)\ d_x\right\}_{s=t} = \int_{x\in\mathbb{X}} x^r e^{tx} f_X(x)\ dx \Longrightarrow M_X^{(r)}(0) = \int_{x\in\mathbb{X}} x^r f_X(x)\ dx = \mathrm{E}_{f_X}[X^r]$$

Hence, in summary, we have that

$$M_X^{(r)}(0) = \mathrm{E}_{f_X}[X^r] \qquad G_X^{(r)}(1) = \mathrm{E}_{f_X}[X(X-1)...(X-r+1)]$$

whenever the functions M_X and G_X exist, and are differentiable, and so the mgf and pgf are potentially useful functions for calculation moments of probability distributions. Also, from the general result for expectations of functions of random variables, if $Y = aX + b$,

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

so that

$$M_Y(t) = e^{bt}M_X(at)$$

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 Cases For $k = 2$, we have that if $Y = X_1 + X_2$ with X_1 and X_2 independent, then

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t)$$

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

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

JOINT PROBABILITY DISTRIBUTIONS

Consider a collection of random variables $X_1, ..., X_k$, with X_i a function from sample space Ω_i to $\mathbb{X}_i \subseteq \mathbb{R}$. Then the **vector** of random variables $\mathbf{X} = (X, ..., X_k)$ is a **vector function** from the Cartesian product of the sample spaces to $\mathbb{X}^{(k)}$

$$\Omega_1 \times ... \times \Omega_k \longrightarrow \mathbb{X}^{(k)} \subseteq \mathbb{X}_1 \times ... \times \mathbb{X}_k \subseteq \mathbb{R}^k$$

The **vector random variable** $\mathbf{X} = (X_1, ..., X_k)$ could correspond to the outcomes of k different experiments carried out once, or a single experiment carried out k times.

In this course we will only consider the case $k = 2$, and the joint probability specification for two variables (X_1, X_2), or (X, Y).

DEFINITION (JOINT MASS FUNCTION)

The **joint probability mass function** of the discrete vector random variable, that is, $\mathbf{X} = (X, Y)$ is denoted $f_{X,Y}$ and is defined by

$$f_{X,Y}(x, y) \;=\; \mathrm{P}[\ (X = x) \cap (Y = y)\ ] \;=\; \mathrm{P}[\ X = x, Y = y\ ]$$

for all possible values of the random vector, $(x, y) \in \mathbb{X} \times \mathbb{Y}$. As for the single variable case, the joint mass function must be **non-negative** for all $\mathbf{x} \in \mathbb{X} \times \mathbb{Y}$, and the sum of the joint mass function evaluated for all $(x, y) \in \mathbb{X} \times \mathbb{Y}$ must be 1.

DEFINITION (MARGINAL MASS FUNCTION)

The **marginal mass function** of random variable X, denoted f_X, is defined in terms of the joint mass function $f_{X,Y}$ for $x \in \mathbb{X}$ by

$$f_X(x) = \sum_{y \in \mathbb{Y}} f_{X,Y}(x, y)$$

that is, the summation of joint mass function evaluated at (x, y) for **all** values of $y \in \mathbb{Y}$. The marginal distribution for random variable Y is defined similarly by summing out over $x \in \mathbb{X}$.

DEFINITION (JOINT PDF)

The **joint probability density function** of continuous random variable (X, Y) is denoted $f_{X,Y}$, and is defined for vector $(x, y) \in \mathbb{X} \times \mathbb{Y}$. It describes the joint probability specification for the two variables. The joint pdf must be non-negative, and integrate to 1 over $(x, y) \in \mathbb{X} \times \mathbb{Y}$.

DEFINITION (MARGINAL PDF)

The **marginal probability density function** of random variable X, denoted f_X, is defined in terms of the joint pdf $f_{X,Y}$ for $x \in \mathbb{X}$ by

$$f_X(x) = \int_{\mathbb{Y}} f_{X,Y}(x, y)\ dy$$

that is, the joint pdf integrated out over the range of the variable Y. The marginal pdf of y defined similarly via an integral with respect to x over the range $\mathbb{X}$.

CONDITIONAL PROBABILITY DISTRIBUTIONS

DEFINITION

The **conditional probability mass/density function** of random variable X, given that $Y = y$ is denoted by

$$f_{X|Y}(x|y)$$

is defined by

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

if the denominator is strictly positive, that is, the ratio of the (full) joint mass function/pdf for the variables X, Y to the marginal mass function/pdf for the variable X evaluated at x. The conditional distribution of Y, given $X = x$ is defined similarly.

Note : This definition is directly related to the definition of **conditional probability** given earlier.

DEFINITION (INDEPENDENCE FOR RANDOM VARIABLES)

Random variables X, Y are **independent** if, for $a_i < b_i$, $i = 1, 2$,

$$\mathrm{P}[\ a_1 < X \leq b_1,\ a_2 < Y \leq b_2\ ] \quad = \quad \mathrm{P}[\ a_1 < X \leq b_1\ ]\mathrm{P}[\ a_2 < Y \leq b_2\ ]$$

In fact, X, Y are independent **if and only if**

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

that is, if and only if the joint mass function/pdf factorizes into the mass function/pdfs. Sufficient conditions for variables X, Y to be independent are

(i) the **support** of the joint mass function/pdf, that is, the region of $\mathbb{R}^2$ on which the joint mass function/pdf is strictly positive, is a Cartesian product.

(ii) the joint mass function/pdf factorizes into a product of marginal functions.

CONDITIONAL EXPECTATION

DEFINITION

Consider two discrete/continuous random variables X and Y with joint mass function/pdf $f_{X,Y}$, and the conditional mass function/pdf of X given $Y = y$, defined in the usual way by

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

Then the **conditional expectation** of X given $Y = y$ is defined by

$$\mathrm{E}_{f_{X|Y}}[X|Y = y] \quad = \quad \begin{cases} \displaystyle\sum_{x \in \mathbb{X}} x f_{X|Y}(x|y) & X \text{ discrete} \\ \\ \displaystyle\int_{\mathbb{X}} x f_{X|Y}(x|y)\ dx & X \text{ continuous} \end{cases}$$

i.e. the expectation of X *with respect to* the conditional density of X given $Y = y$, (possibly a function of y).

BIVARIATE EXPECTATIONS

For random variables X, Y mass/density function $f_{X,Y}$, the **expectation** or **expected value** of function $g(X,Y)$ of X and Y is defined by

$$\mathrm{E}_{f_{X,Y}}[g(X,Y)] = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) \\ \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\ dxdy \end{cases}$$

in the discrete and continuous case respectively.

Properties

(i) Let X, Y be random variables with joint mass function/pdf $f_{X,Y}$. Let g_1 and g_2 be real-valued functions and a and b be constants. Then

$$\mathrm{E}_{f_{X,Y}}[ag_1(X,Y) + bg_2(X,Y)] = a\mathrm{E}_{f_{X,Y}}[g_1(X,Y)] + b\mathrm{E}_{f_{X,Y}}[g_2(X,Y)]$$

(ii) Let X, Y be independent random variables with mass functions/pdfs f_X and f_Y respectively. Let g_1, g_2 be scalar functions. Then if

$$g(X,Y) \ = \ g_1(X)g_2(Y)$$

then

$$\mathrm{E}_{f_{X,Y}}[g(X,Y)] \ = \ \mathrm{E}_{f_X}[g_1(X)]\mathrm{E}_{f_Y}[g_2(Y)]$$

THE LAW OF ITERATED EXPECTATION

For two continuous random variables X and Y with joint pdf $f_{X,Y}$,

$$\begin{aligned}
\mathrm{E}_{f_X}[X] &= \int_{\mathbb{X}} x f_X(x)\ dx = \int_{\mathbb{X}} x \left\{\int_{\mathbb{Y}} f_{X,Y}(x,y)\ dy\right\}\ dx \\
&= \int_{\mathbb{X}} x \left\{\int_{\mathbb{Y}} f_{X|Y}(x|y) f_Y(y)\ dy\right\} \\
&= \int_{\mathbb{X}}\int_{\mathbb{Y}} x f_{X|Y}(x|y) f_Y(y)\ dy\ dx \\
&= \int_{\mathbb{Y}} \left\{\int_{\mathbb{X}} x f_{X|Y}(x|y)\ dx\right\} f_Y(y)\ dy \\
&= \int_{\mathbb{Y}} \left\{\mathrm{E}_{f_{X|Y}}[X|Y=y]\right\} f_Y(y)\ dy = \mathrm{E}_{f_Y}\left[\mathrm{E}_{f_{X|Y}}[X|Y=y]\right]
\end{aligned}$$

so the expectation of X can be calculated by finding the **conditional expectation** of X given $Y = y$, giving a function of y, and then taking the expectation of this function with respect to the marginal density for Y. Note that this proof only works if the conditional expectation and the marginal expectation are finite.

COVARIANCE OF TWO RANDOM VARIABLES

In the single variable case, it was possible to compute moments and central moments for a probability distribution; in particular, the variance of the distribution, defined as

$$\mathrm{Var}_{f_X}[X] = \mathrm{E}_{f_X}[(X-\mu)^2]$$

where $\mu = \mathrm{E}_{f_X}[X]$, so that the variance is the second central moment of the distribution. The variance quantifies the amount of *spread* in the distribution. This idea can be extended to the two variable case:

DEFINITION

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

$$\mathrm{Cov}_{f_{X,Y}}[X,Y] = \mathrm{E}_{f_{X,Y}}[(X-\mu_X)(Y-\mu_Y)]$$

where μ_X and μ_Y are the marginal expectations of X and Y respectively.

Writing out the expectation in full, by expanding the function $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$,

$$\mathrm{Cov}_{f_{X,Y}}[X,Y] = \mathrm{E}_{f_{X,Y}}[XY] - \mu_X\mu_Y = \mathrm{E}_{f_{X,Y}}[XY] - \mathrm{E}_{f_X}[X]\ \mathrm{E}_{f_Y}[Y]$$

DEFINITION

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

$$\mathrm{Corr}_{f_{X,Y}}[X,Y] = \frac{\mathrm{Cov}_{f_{X,Y}}[X,Y]}{\sqrt{\mathrm{Var}_{f_X}[X]\ \mathrm{Var}_{f_Y}[Y]}}$$

If $\mathrm{Cov}_{f_{X,Y}}[X,Y] = \mathrm{Corr}_{f_{X,Y}}[X,Y] = 0$ then variables X and Y are *uncorrelated.*

THEOREM

If random variables X and Y are *independent* then $\mathrm{Cov}_{f_{X,Y}}[X,Y] = 0$, and X and Y are uncorrelated.

Proof

If X and Y are independent, then it is easy to see that

$$\mathrm{E}_{f_{X,Y}}[XY] = \mathrm{E}_{f_X}[X]\mathrm{E}_{f_Y}[Y]$$

and hence the result follows from the covariance definition.

NOTE:
For random variables X and Y, with (marginal) expectations μ_X and μ_Y respectively, and (marginal) variances σ_X^2 and σ_Y^2 respectively, if random variables Z_1 and Z_2 are defined

$$\begin{aligned} Z_1 &= (X-\mu_X)/\sigma_X \\ Z_2 &= (Y-\mu_Y)/\sigma_Y \end{aligned}$$

that is, Z_1 and Z_2 are *standardized* variables, then

$$\mathrm{Corr}_{f_{X,Y}}[X,Y] = \mathrm{Cov}_{f_{Z_1,Z_2}}[Z_1,Z_2]$$

BIVARIATE TRANSFORMATIONS OF RANDOM VARIABLES

As in the single variable case, it is often possible to consider and calculate the probability distribution of transformed random variables; for example, given the joint probability distribution of random variables X and Y, it is often possible to calculate the joint probability distribution of, say, $U = g_1(X, Y)$ and $V = g_2(X, Y)$ for general functions g_1 and g_2. The techniques required for such calculations are reasonably straightforward, and analogous to the single variable case, but beyond the scope of this course. Here, we look only at a special case concerning sums of random variables.

SUMS OF RANDOM VARIABLES

It is often necessary to calculate the probability distribution of a sum of random variables. The following theorem outlines the technique via which such a calculation can be carried out for the sum of two random variables.

THEOREM - THE CONVOLUTION FORMULA

Let X and Y be random variables respectively with joint mass/density function $f_{X,Y}$. Then, if random variable T is defined by $T = X + Y$, the mass/density of T is given by

$$f_T(t) = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} f_{X,Y}(x, t-x) & \text{DISCRETE} \\ \\ \displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x, t-x)\ dx & \text{CONTINUOUS} \end{cases}$$

If X and Y are **independent**, then

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

so the summand/integrand factorizes.

This result is important as it allows routine calculation of the distribution of sums of random variables. In the discrete case, this result is merely a version of the Theorem of Total Probability, as

$$(T = t) = \bigcup_{x=-\infty}^{\infty} ((X = x) \cap (Y = t - x))$$

$$\Longrightarrow \quad \mathrm{P}[\ T = t\ ] = \sum_{x=-\infty}^{\infty} \mathrm{P}[\ X - x,\ Y = t - x\ ]$$

as the events $\{(X = x) \cap (Y = t - x)\}$ are mutually exclusive for different values of x.

EXPECTATION RESULTS FOR SUMS OF RANDOM VARIABLES

Results for linear combinations of random variables: as a special case of the above result, if random variable X is defined by

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

for random variables X, Y and constants $a_1, ..., a_k$, then

$$\mathrm{E}_{f_X}[X] = \sum_{i=1}^{k} a_i \mathrm{E}_{f_{X_i}}[X_i]$$

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

Note also that, if Z_1 and Z_2 are the standardized variables, defined in the previous section,

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

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

THE CENTRAL LIMIT THEOREM

Suppose $X_1, ..., X_n$ are independent and identically distributed random variables with $\mathrm{E}_{f_X}[X_i] = \mu$ and $\mathrm{Var}_{f_X}[X_i] = \sigma^2$ that is, the expectation and variance of the X_is are specified, but the distribution is unspecified. Let the random variable Z_n be defined by

$$Z_n = \frac{\displaystyle\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

Then as $n \longrightarrow \infty$, the distribution of Z_n tends to a standard normal distribution.

This result is important as the representation of a random variable as the sum of independent and identically distributed random variables occurs naturally in many probability models; for example, the Binomial and Poisson distributions arise as sums of independent and identically distributed Bernoulli random variables; the Negative Binomial arises as the sum of independent and identically distributed Geometric random variables; the Gamma distribution arises as the sum of independent and identically Exponential random variables. The Central Limit Theorem result indicates that, when the number of terms in the sum tends to infinity, the probability distribution of the sum tends to the standard normal, and hence it is possible to construct approximations to based on the normal distribution to facilitate probability calculations.

APPENDIX: EXPECTATION AND GENERATING FUNCTION RESULTS FOR STANDARD DISTRIBUTIONS

DISCRETE PROBABILITY DISTRIBUTIONS

BERNOULLI DISTRIBUTION

MASS FUNCTION

$$f_X(x) = \theta^x (1-\theta)^{1-x} \qquad \text{for } x \in \{0,1\} \qquad \text{where } 0 \leq \theta \leq 1.$$

MGF

$$M_X(t) \;=\; \sum_{x=0}^{1} \mathrm{e}^{tx}\theta^x(1-\theta)^{1-x} \;=1-\theta+\theta\mathrm{e}^t$$

rth MOMENT

$$M_X^{(r)}(t) = \theta\mathrm{e}^t \Longrightarrow M_X^{(r)}(0) = \theta \qquad r = 1,2,...$$

$$\Longrightarrow \mathrm{E}_{f_X}[X] = \theta \qquad \mathrm{E}_{f_X}[X^2] = \theta \Longrightarrow \mathrm{Var}_{f_X}[X] = \;\theta - \theta^2 \;=\; \theta(1-\theta)$$

BINOMIAL DISTRIBUTION

MASS FUNCTION

$$f_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad \text{for } x \in \{0,1,2,...,n\} \qquad \text{where } n \geq 0,\; 0 \leq \theta \leq 1.$$

MGF

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} \theta^x (1-\theta)^{n-x} \;=\; \sum_{x=0}^{n} \binom{n}{x} \left(\theta e^t\right)^x (1-\theta)^{n-x} = \left(1-\theta+\theta e^t\right)^n$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = n\theta\mathrm{e}^t\left(1-\theta+\theta\mathrm{e}^t\right)^{n-1} \qquad M_X^{(2)}(t) = n(n-1)\left\{\theta\mathrm{e}^t\right\}^2\left(1-\theta+\theta\mathrm{e}^t\right)^{n-2} + n\theta\mathrm{e}^t\left(1-\theta+\theta\mathrm{e}^t\right)^{n-1}$$

so that $M_X^{(1)}(0) = n\theta$ and $M_X^{(2)}(0) = n(n-1)\theta^2 + n\theta$, and thus

$$\mathrm{E}_{f_X}[X] = \; n\theta \qquad \mathrm{Var}_{f_X}[X] = \; n(n-1)\theta^2 + n\theta - n^2\theta^2 = n\theta(1-\theta)$$

POISSON DISTRIBUTION

MASS FUNCTION

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x \in \{0,1,2,...\} \qquad \text{where } \lambda > 0.$$

MGF

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} \; = \; e^{-\lambda}\sum_{x=0}^{\infty}\frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda}e^{\lambda e^t} \; = \; \exp\left\{\lambda\left(e^t - 1\right)\right\}$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \; \lambda e^t \exp\left\{\lambda\left(e^t - 1\right)\right\} \qquad M_X^{(2)}(t) = \left(\lambda e^t\right)^2 \exp\left\{\lambda\left(e^t - 1\right)\right\} + \lambda e^t \exp\left\{\lambda\left(e^t - 1\right)\right\}$$

so that $M_X^{(1)}(0) = \lambda$ and $M_X^{(2)}(0) = \lambda^2 + \lambda$, and thus

$$\mathrm{E}_{f_X}[X] = \; \lambda \qquad \mathrm{Var}_{f_X}[X] = \; \lambda^2 + \lambda - \lambda^2 = \lambda$$

GEOMETRIC DISTRIBUTION

MASS FUNCTION

$$f_X(x) = (1-\theta)^{x-1}\theta \qquad \text{for } x \in \{1,2,...\} \qquad \text{where } 0 \le \theta \le 1.$$

MGF

$$M_X(t) \; = \; \sum_{x=1}^{\infty} e^{tx}(1-\theta)^{x-1}\theta \; = \; \theta e^t \sum_{x=1}^{\infty} e^{t(x-1)}(1-\theta)^{x-1} \; = \; \theta e^t \sum_{x=0}^{\infty}\left(e^t(1-\theta)\right)^x \; = \; \frac{\theta e^t}{1 - e^t(1-\theta)}$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \; \frac{\theta e^t}{\left[1 - e^t(1-\theta)\right]^2} \qquad M_X^{(2)}(t) = \; \frac{\theta e^t\left[1 - e^t(1-\theta)\right]\left[1 + e^t(1-\theta)\right]}{\left[1 - e^t(1-\theta)\right]^4}$$

so that $M_X^{(1)}(0) = \dfrac{1}{\theta}$ and $M_X^{(2)}(0) = \dfrac{2-\theta}{\theta^2}$, and thus

$$\mathrm{E}_{f_X}[X] = \; \frac{1}{\theta} \qquad \mathrm{Var}_{f_X}[X] = \; \frac{2-\theta}{\theta^2} - \frac{1}{\theta^2} \; = \; \frac{1-\theta}{\theta^2}$$

NEGATIVE BINOMIAL DISTRIBUTION

MASS FUNCTION

$$f_X(x) = \begin{pmatrix} x-1 \\ n-1 \end{pmatrix} \theta^n (1-\theta)^{x-n} \qquad \text{where } n \in \{1,2,3,...\}\,,\ 0 \le \theta \le 1.$$

MGF

$$M_X(t) = \sum_{x=n}^{\infty} \mathrm{e}^{tx} \begin{pmatrix} x-1 \\ n-1 \end{pmatrix} \theta^n (1-\theta)^{x-n} \;=\; \left(\theta \mathrm{e}^t\right)^n \sum_{x=n}^{\infty} \begin{pmatrix} x-1 \\ n-1 \end{pmatrix} \left(\mathrm{e}^t(1-\theta)\right)^{x-n} = \left\{ \frac{\theta \mathrm{e}^t}{1-\mathrm{e}^t(1-\theta)} \right\}^n$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \frac{n(\theta \mathrm{e}^t)^n}{[1-\mathrm{e}^t(1-\theta)]^{n+1}} \qquad M_X^{(2)}(t) = \frac{n(\theta \mathrm{e}^t)^n \left[n + \mathrm{e}^t(1-\theta)\right]}{[1-\mathrm{e}^t(1-\theta)]^{n+2}}$$

so that $M_X^{(1)}(0) = \dfrac{n}{\theta}$ and $M_X^{(2)}(0) = \dfrac{n(n+(1-\theta))}{\theta^2}$, and thus

$$\mathrm{E}_{f_X}[X] = \frac{n}{\theta} \qquad \mathrm{Var}_{f_X}[X] = \frac{n(n+(1-\theta))}{\theta^2} - \frac{n^2}{\theta^2} = \frac{n(1-\theta)}{\theta^2}$$

CONTINUOUS PROBABILITY DISTRIBUTIONS

CONTINUOUS UNIFORM DISTRIBUTION

PDF

$$f_X(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

rth MOMENT

$$\mathrm{E}_{f_X}[X^r] \;=\; \int_a^b x^r \frac{1}{b-a}\, dx \;=\; \frac{1}{b-a}\left[\frac{b^{r+1}}{r+1} - \frac{a^{r+1}}{r+1}\right]$$

so therefore

$$\begin{aligned}
\mathrm{E}_{f_X}[X] \;&=\frac{1}{b-a}\left[\frac{b^2-a^2}{2}\right] \;=\; \frac{(a+b)}{2}\\
\mathrm{E}_{f_X}[X^2] \;&=\frac{1}{b-a}\left[\frac{b^3-a^3}{3}\right] \;=\; \frac{(a^2+ab+b^2)}{3}\\
\Longrightarrow \mathrm{Var}_{f_X}[X] \;&=\frac{(a^2+ab+b^2)}{3} - \left\{\frac{(a+b)}{2}\right\}^2\\
&=\frac{(a-b)^2}{12}
\end{aligned}$$

EXPONENTIAL DISTRIBUTION

PDF

$$f_X(x) = \lambda \mathrm{e}^{-\lambda x} \qquad x > 0 \qquad \text{where } \lambda > 0.$$

MGF

$$M_X(t) \;=\; \int_0^\infty \mathrm{e}^{tx}\lambda\mathrm{e}^{-\lambda x}\, dx = \lambda\int_0^\infty \mathrm{e}^{-(\lambda-t)x}\, dx = \frac{\lambda}{\lambda-t} \qquad \text{for } t < \lambda$$

rth MOMENT

$$M_X^{(r)}(t) = \frac{r!\lambda}{(\lambda-t)^{r+1}} \Longrightarrow M_X^{(r)}(0) = \frac{r!}{\lambda^r}$$

and therefore, evaluating for $r = 1$ and 2, we have

$$\mathrm{E}_{f_X}[X] = \frac{1}{\lambda},\; \mathrm{E}_{f_X}[X^2] = \frac{2}{\lambda^2} \Longrightarrow \mathrm{Var}_{f_X}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \;=\; \frac{1}{\lambda^2}$$

GAMMA DISTRIBUTION

PDF

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x} \qquad x > 0 \qquad \text{where } \alpha, \beta > 0.$$

MGF

$$M_X(t) = \int_0^\infty \mathrm{e}^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}\, dx \;=\; \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1}\, \mathrm{e}^{-(\beta-t)x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\, \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} = \left(\frac{\beta}{\beta-t}\right)^\alpha$$

as the integrand is a pdf (of a $Gamma(\alpha, \beta - t)$ random variable).

rth MOMENT

$$M_X^{(1)}(t) = \frac{\alpha\beta^\alpha}{(\beta-t)^{\alpha+1}} \quad M_X^{(2)}(t) = \frac{\alpha(\alpha+1)\beta^\alpha}{(\beta-t)^{\alpha+2}} \Longrightarrow M_X^{(1)}(0) = \frac{\alpha}{\beta} \qquad M_X^{(2)}(0) = \frac{\alpha(\alpha+1)}{\beta^2}$$

and thus

$$\mathrm{E}_{f_X}[X] = \frac{\alpha}{\beta} \quad \mathrm{Var}_{f_X}[X] = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$$

STANDARD NORMAL DISTRIBUTION

PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} \qquad x \in \mathbb{R}$$

MGF

$$\begin{aligned} M_X(t) &= \int_{-\infty}^\infty \mathrm{e}^{tx} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{-\frac{1}{2}x^2 + tx\right\}\, dx \\ &= \exp\left\{\frac{t^2}{2}\right\} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-t)^2\right\} dx = \exp\left\{\frac{t^2}{2}\right\} \end{aligned}$$

by first completing the square in x, and then by noting that the integrand is a pdf (of an $N(t,1)$ variable), and thus the integral is equal to one.

Hence we can write down the mgf of a non-standard normal random variable. If $X \sim N(0,1)$, then $Y = \sigma X + \mu \;\sim N(\mu, \sigma^2)$ from previous transformation results. Hence by the standard mgf result, we have

$$M_Y(t) = \mathrm{E}_{f_Y}[e^{tY}] = \mathrm{E}_{f_X}\left[e^{(\sigma X + \mu)t}\right] = e^{\mu t}\mathrm{E}_{f_X}\left[e^{\sigma t X}\right] = e^{\mu t} M_X(\sigma t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = (\mu + t\sigma^2)\exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \qquad M_X^{(2)}(t) = (\mu^2 + 2t\sigma^2\mu + t^2\sigma^4 + \sigma^2)\exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\}$$

so that $M_X^{(1)}(0) = \mu$ and $M_X^{(2)}(0) = \mu^2 + \sigma^2$, and thus

$$\mathrm{E}_{f_X}[X] = \mu \qquad \mathrm{Var}_{f_X}[X] = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

EXPECTATIONS AND VARIANCES OF STANDARD DISTRIBUTIONS

	Parameters	EXPECTATION	VARIANCE
Discrete Distributions			
$Bernoulli(\theta)$	θ	θ	$\theta(1-\theta)$
$Binomial(n,\theta)$	n,θ	$n\theta$	$n\theta(1-\theta)$
$Poisson(\lambda)$	λ	λ	λ
$Geometric(\theta)$	θ	$\frac{1}{\theta}$	$\frac{(1-\theta)}{\theta^2}$
$NegBinomial(n,\theta)$	n,θ	$\frac{n}{\theta}$	$\frac{n(1-\theta)}{\theta^2}$
$Hypergeometric(N,R,n)$	N,R,n	$\frac{nR}{N}$	$\frac{nR(N-R)(N-n)}{N^2(N-1)}$
Continuous Distributions			
$Uniform(a,b)$	a,b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Exponential(\lambda)$	λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$Gamma(\alpha,\beta)$	α,β	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
$Normal(\mu,\sigma^2)$	μ,σ^2	μ	σ^2

MOMENT GENERATING FUNCTIONS OF STANDARD DISTRIBUTIONS

	Parameters	MGF
Discrete Distributions		
$Bernoulli(\theta)$	θ	$1-\theta+\theta \mathrm{e}^t$
$Binomial(n,\theta)$	n,θ	$(1-\theta+\theta \mathrm{e}^t)^n$
$Poisson(\lambda)$	λ	$\exp\left\{\lambda(e^t-1)\right\}$
$Geometric(\theta)$	θ	$\dfrac{\theta \mathrm{e}^t}{1-\mathrm{e}^t(1-\theta)}$
$NegBinomial(n,\theta)$	n,θ	$\left\{\dfrac{\theta \mathrm{e}^t}{1-\mathrm{e}^t(1-\theta)}\right\}^n$
Continuous Distributions		
$Exponential(\lambda)$	λ	$\dfrac{\lambda}{\lambda-t}$
$Gamma(\alpha,\beta)$	α,β	$\left(\dfrac{\beta}{\beta-t}\right)^\alpha$
$Normal(\mu,\sigma^2)$	μ,σ^2	$\exp\left\{\mu t+\dfrac{\sigma^2 t^2}{2}\right\}$