ECE 6010
Lecture 2 – More on Random Variables

Readings from G&S: Section 3.3, Section 3.4, Section 3.6, Section 3.7, Section 4.3, Section 4.6, Section 5.1, Section 5.2, Section 5.6, Section 5.7, Section 5.8,

Expectation

When we say "expectation," we mean "average," the average being roughly what you would think of (i.e., the arithmetic average, as opposed to a median or mode). For a discrete r.v. X, we define the **expectation** as

$$\boxed{E[X] = \sum_i x_i p_X(x_i)}$$

For a continuous r.v., we define the expectation as

$$\boxed{E[X] = \int_{-\infty}^{\infty} f_X(x) x \, dx}$$

Now a bit of technicality regarding integration, which introduces notation commonly used. When you integrate, you are typically doing a Riemann integral:

$$\int_a^b x f_X(x) \, dx = \lim_{\max_{1 \le i \le n-1} |x_{i+1} - x_i| \to 0} \sum_{i=1}^n z_i f_X(z_i)(x_{i+1} - x_i)$$

$$a = x_1 < x_2 < \cdots < x_n = b \qquad\qquad z_i \in (x_i, x_{i+1})$$

In other words, we break up the interval into little slices and add up the vertical rectangular pieces.

Another way of writing this is to recognize that

$$z_i f_X(z_i)(x_{i+1} - x_i) \approx z_i P(x_i < X \le X_{i+1}) = z_i(F_X(x_{i+1}) - F_X(x_i))$$

and that in the limit, the approximation becomes exact. Note, however, that this is expressed in terms of the c.d.f., not the p.d.f., and so exists for all random variables, not just continuous ones.

This gives rise to what is known as the Riemann-Stieltjes Integral:

$$E[X] = \lim_{\max_{1 \le i \le n-1}(x_{i+1} - x_i) \to 0} \sum_{i=1}^n z_i [F_X(x_{i+1}) - F_X(x_i)].$$

We write the limit as

$$\int_a^b x dF_X(x)$$

This notation "describes" continuous, discrete, and mixed cases. That is,

$$E[X] = \int_{-\infty}^{\infty} x dF_X(x).$$

We have defined the Riemann-Stieltjes integral in a context of expectation. However, it has a more general definition:

$$\int_a^b f(x) dg(x) = \lim_{\max(x_{i+1} - x_i) \to 0} \sum_i f(z_i)[g(x_{i+1}) - g(x_i)]$$

When $g(x) = x$, this reduces to the ordinary Riemann integral. Sufficient conditions for existence:

- $g(x)$ of bounded variation
- and $f(x)$ continuous on $[a, b]$

or

- $f(x)$ of bounded variation
- $g(x)$ continuous

The first case covers the case of expectation.

In a directly analogous way we define

$$\int_{-\infty}^{\infty} g(x)dF_X(x) = \lim \sum_{i=1}^{n} g(z_i)[F_X(x_{i+1}) - F_X(x_i)].$$

Now consider the r.v. $Y = g(X)$.

$$E[Y] = \int_{-\infty}^{\infty} y dF_Y(y)$$

Note that $dF_Y(y)$ is the representation of the limiting value

$$\begin{aligned} F_Y(y_{i+1}) - F_Y(y_i) &= Pr(y_i < Y \leq y_{i+1}) = Pr(y_i < g(X) \leq y_{i+1}) \\ &= Pr(g^{-1}(y_i) < X \leq g^{-1}(y_{i+1})) = Pr(x_i < X \leq x_{i+1}) \end{aligned}$$

which, in the limit is equal to $dF_X(x)$, when $y = g(x)$. Thus

$$\boxed{E[Y] = \int_{-\infty}^{\infty} y dF_Y(y) = \int_{\infty}^{\infty} g(x) dF_X(x)}$$

Let us put this in more familiar terms: If $Y = g(x)$, then

$$\boxed{E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx} \tag{1}$$

One might think that finding $E[Y]$ would require finding $f_Y(y)$. However, as (1) shows, all that is necessary is to substitute $g(x)$ for x in the expectation. This is sometimes called the *law of the unconcious statistitian*, since it can be done nearly thoughtlessly.

An interesting result is obtained through the use of **indicator** functions. Let $I : (\mathcal{F}, \Omega) \to \mathbb{R}$ be defined by

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

In other words, the indicator function indicates which its argument is in the set which is the subscripted argument.

We define a **simple function** as one which is a linear combination of indicator functions: For some collection $A_1, A_2, \ldots, A_n \in \mathcal{F}$,

$$g(\omega) = \sum_{k=1}^{n} b_k I_{A_k}(\omega)$$

This gives us a piecewise-constant function on Ω. It also defines a random variable.

Note that the collection need not be disjoint. However, we can shuffle things around to write the function as

$$g(\omega) = \sum_{k=1}^{n^*} b_k^* I_{A_k^*}(\omega)$$

where the A_k^*s *are* disjoint, and where the b_k^*s are unique. Note that

$$A_k^* = \{\omega \in \Omega : g(\omega) = b_k^*\}$$

Now note that

$$\boxed{E[I_A] = P(A)}$$

Based on this, and the disjointness of the A_k^*, we can write

$$E[g] = \sum_{i=1}^{n^*} b_k^* P(A_k^*)$$

There are many instances where indicator functions are used to get a "handle" on the probability of an event.

Now we will get a bit more technical, dealing with some issues related to the existence of expectations. We have seen how to define expectations for simple functions (which are random variables). But what about more general random variables? Let $X \geq 0$ be a random variable. We *define*

$$E[X] = \sup_{g \text{simple} \leq X} E[g]$$

where the $\sup$ means "least upper bound", and the limit is taken over all simple functions g satisfying $g \leq X$. It can be shown that this limit always exists, though it may be infinite. There is thus no question of convergence or anything like that.

Generalizing further, let X be an arbitrary r.v. Since the previous result holds for non-negative random variables, let us split X:

$$X = X^+ - X^-$$

where X^+ is the positive part and X^- is the negative part:

$$X^+ = \max(X(\omega), 0) \geq 0 \qquad\qquad X^- = -\min(X(\omega), 0) \geq 0$$

Now X^+ and X^- have well-defined expectations. We take

$$E[X] = E[X^+] - E[X^-]$$

which is defined in every case except when $E[X^+]$ and $E[X^-]$ are both infinite (in which case the difference is undefined).

Let us examine the expectation in light of the Riemann-Stieltjes integral. We define

$$E[X] = \int_{-\infty}^{\infty} x dF_X(x) = \lim_{a \to -\infty, b \to \infty} \int_a^b x dF_X(x)$$

This is a stronger sense of the limit than, for example

$$\int_{-\infty}^{\infty} x dF_X(x) = \lim_{a \to \infty} \int_{-a}^{a} x dF_X(x)$$

For example, $\sin(x)$ has an integral in the latter sense (which is equal to 0), but not in the former sense.

Now we will consider an example of a density where the expectation does not exist.

Example 1 Cauchy density:

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

It is straightforward to show that this satisfies the requirements for a p.d.f. It looks a lot like a Gaussian, but has "heavy tails." It can be shown that a Cauchy r.v. can be obtained as a ratio of Gaussians: $X = Y/Z$. Now let us attempt to compute

$$E[X] = \lim_{a \to -\infty, b \to \infty} \frac{1}{\pi} \int_a^b \frac{x}{1+x^2}\, dx$$

If $a = -b$, then the integral is zero, no matter what. If you fix a or b, taking the limit of the other one, the result is ∞. That is, $E[X^+]$ exists and $E[X^-]$ exists (although both are ∞), but they can't be subtracted. □

Properties of Expectations

1. If $X = c$ then $E[X] = c$.
2. If $Y = aX + b$ then $E[Y] = aE[X] + b$.

$E[X]$ acts kind of like an integral of $X(\omega)$ over Ω, weighted by P. One way that the expectation is expressed is

$$E[X] = \int_\Omega X(\omega)P(d\omega) = \int_\Omega X\,dP.$$

An integral in this form is said to be a Lebesgue-Stieltjes Integral. Since X induces a probability P_X on $(\mathbb{R}, \mathcal{B})$, as we have observed we can also think of the probability space $(\mathbb{R}, \mathcal{B}, P_X)$. We can write

$$E[X] = \int_\mathbb{R} x P_X(dx)$$

where now X is the "identity" r.v. on the real line. We thus have two equivalent definitions:

$$\int_\Omega X(\omega)P(d\omega) = \int_\mathbb{R} x P_X(dx)$$

Back to properties:

1. If $Y = g \circ X$ then

$$E[Y] = \int_\Omega (g \circ X)(\omega)P(d\omega) = \int_\mathbb{R} g(x)P_X(dx) = \int_\mathbb{R} y P_Y(dy)$$

Pairs of random variables

Ultimately, we will be dealing with infinite sequences of random variables. As steps along the way, we will examine carefully pairs of random variables, then vectors of random variables.

On $\mathbb{R}^2$, the smallest σ-field of interest is $\mathcal{B}^2$, which is the smallest σ-field containing all of the rectangles. This is the Borel σ-field of $\mathbb{R}^2$.

Definition 1 A **bivariate random variable** (X, Y) is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^2, \mathcal{B}^2)$. □

That is,

$$\{\omega \in \Omega : (X, Y)(\omega) \in B\} \in \mathcal{F} \forall B \in \mathcal{B}^2$$

Note that two r.v.s X, Y on $(\Omega, \mathcal{F})$ form a bivariate r.v.

Definition 2 The **joint** or **bivariate** distribution of (X, Y) is

$$P_{XY}(B) = P(\{\omega \in \Omega : (X, Y)(\omega) \in B\})$$

for $B \in \mathcal{B}^2$. □

Definition 3 The **joint c.d.f.** of (X, Y) is defined as

$$F_{XY}(a, b) = P(X \leq a, Y \leq b) = P((X, Y) \in R_{a,b})$$

where R_{ab} is the semi-infinite rectangle

$$R_{a,b} = \{(x, y) \in \mathbb{R}^2 : x \leq a, y \leq b\}.$$

□

Properties of the joint c.d.f.:

1. $\lim_{a,b\to\infty} F_{X,Y}(a, b) = 1$.
2. $\lim_{a\to-\infty} F_{X,Y}(a, b) = 0 = \lim_{b\to-\infty} F_{X,Y}(a, b)$.
3. $\lim_{a\to\infty} F_{X,Y}(a, b) = F_Y(b)$, the marginal c.d.f. of Y.
   $\lim_{b\to\infty} F_{X,Y}(a, b) = F_X(a)$, the marginal c.d.f. of X.
4. $F_{XY}(a, b)$ is continuous "from the northeast."
5. $F_{XY}(x, y)$ is montonically increasing (or, more precisely, nondecreasing) in both variables.

Any function with these properties is a legitimate c.d.f., and completely characterizes the family of joint c.d.f.s.

Joint discrete r.v.s

Definition 4 If X, Y are discrete r.v.s taking values in sets $\{x_1, \ldots\}$ and $\{y_1, \ldots, \}$, respectively, then (X, Y) forms a discrete bivariate r.v. and its joint p.m.f. is defined by

$$p_{XY}(a, b) = P(X = a, Y = b)$$

□

Properties of p_{XY}:

1. $p_{XY} \geq 0$, and $p_{XY}(a, b) = 0$ if $a \notin \{x_1, \ldots\}$ or $b \notin \{y_1, \ldots\}$
2. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{XY}(x_i, y_j) = 1$.
3. $F_{XY}(a, b) = \sum_{\{x_i, y_i\}: x_i \leq a, y_j \leq b\}} p_{XY}(x_i, y_j)$
4. Marginals:

$$P_X(x_i) = \sum_j p_{XY}(x_i, y_j)$$

$$P_Y(y_j) = \sum_i p_{XY}(x_i, y_j)$$

Joint continuous r.v.s

Definition 5 X and Y are **jointly continuous** r.v.s if there is a function $f_{XY} : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$F_{XY}(a, b) = \int_{\infty}^{b} \int_{\infty}^{a} f_{XY}(x, y)\, dxdy$$

for all $(a, b) \in \mathbb{R}^2$. The function f_{XY} is called the **joint p.d.f.** of X and Y (when it exists).

□

Properties of joint p.d.f.:

1. $f_{XY} \geq 0$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx, dy = 1$.

3. We can get the p.d.f. from the c.d.f:

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y).$$

4. Marginals:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

If X and J are jointly continuous, then they are marginally continuous (that is, f_X and f_Y are continuous). However, the opposite is not true.

Example 2 Important example! X and Y are said to be **jointly Gaussian** with parameters μ_x, μ_y and σ_x^2, σ_y^2, (where $\mu_x, \mu_y \in \mathbb{R}$ and $\sigma_x^2, \sigma_y^2 \geq 0$) and ρ (where $\rho \in (-1, 1)$) if the have the joint p.d.f.

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

We write $(X,Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$. In this case, the marginals satisfy

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2) \qquad\qquad Y \sim \mathcal{N}(\mu_y, \sigma_y^2).$$

If we make contour plots of the function, that is, plots of constant probability, we obtain ellipses. When $\rho = 0$, we get circles. As ρ increases positive, the ellipses tilt clockwise. As ρ decreases, the ellipses tilt counterclockwise.

ρ is called the correlation coefficient. It is a measure of how much X “looks like” Y.

$\square$

Independence of r.v.s

Definition 6 X and Y are **independent** if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \forall A, B \in \mathcal{B}.$$

$\square$

1. X and Y are independent iff $F_{XY}(a,b) = F_X(a)F_Y(b) \forall (a,b) \in \mathbb{R}^2$.
2. if X and Y are jointly continuous, then they are independent iff $f_{XY}(x,y) = f_X(x)f_Y(y)$
3. If X and Y are discrete, then they are independent iff $p_{XY}(a,b) = p_X(a)p_Y(b)$.
4. If X and Y are jointly Gaussian random variables, then they are independent iff $\rho = 0$. (Show this using the p.d.f.)

Caution: Gaussian r.v.s are special this way. As a general rule, **uncorrelated does not imply independence.**.

In practice, it is common to assume that random variables are independent based on physical arguments, rather than to prove is by identifying a joint density and computing the marginals.

Many times, independence is also taken as an assumption, even when it is not strictly true. This independence assumption frequently simplifies analysis. However, the validity of the assumption must be validated (e.g., using computer simulations).

Expectations of functions of two r.v.s

Let $g : \mathbb{R}^2 \to \mathbb{R}$ be measurable (e.g., $\{(x, y) \in \mathbb{R}^2 : g(x, y) \in B\} \in \mathcal{B}^2 \forall B \in \mathcal{B}^2$). Then for a bivariate r.v. (X, Y) we can define $Z = g(X, Y)$.

$$E[Z] = \int_{-\infty}^{\infty} z dF_Z(z) = \int_{\mathbb{R}} g(x, y) P_{XY}(dx\, dxy)$$
$$= \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dxdy & (X, Y) \text{ jointly continuous} \\ \sum_i \sum_j g(x_i, y_j) p_{XY}(x_i, y_j) & (X, Y) \text{ discrete} \end{cases}$$

Properties:

1. $E[X + Y] = E[X] + E[Y]$
2. If $X \geq Y$ then $E[X] \geq E[Y]$.
3. If X and Y are independent then

   $$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)] \forall (\text{measurable, well-defined}) g_1, g_2.$$

   Comments: If X and Y are independent, then

   $$E[XY] = E[X]E[Y].$$

   However, if $E[X]E[Y] = E[XY]$, this does *not* mean that they are independent. (Uncorrelated does not imply independence.)

   However, if $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$ for *all* appropriate functions, then X and Y are independent. In fact, this is necessary and sufficient for independence.

Definition 7 The **covariance** of X and Y are is defined as

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

The **variance** of X is defined as

$$\text{var}(X) = \text{cov}(X, X).$$

□

4. $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$. $\qquad \text{var}(X) = E[X^2] - (E[X])^2$.
5. If X and Y are independent then $\text{cov}(X, Y) = 0$. If $\text{cov}(X, Y) = 0$, we say that X and Y are **uncorrelated.**

   Again, uncorrelated does not imply independence.
6. $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X, Y)$. If $\text{cov}(X, Y) = 0$ then $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.
7. $\text{cov}(aX + b, cY + d) = ac\,\text{cov}(X, Y)$ for all constants $a, b, c, d \in \mathbb{R}$. Thus

   $$\text{var}(aX) = a^2\,\text{var}(X).$$

Definition 8 If $0 < \text{var}(X) < \infty$ and $0 < \text{var}(Y) < \infty$, the **correlation coeffiicent** between X and Y is

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}}.$$

This is a normalized version of the covariance. □

8. $|\rho| \leq 1$. This can be shown using the Cauchy-Schwartz inequality.
   $|\rho| = 1$ iff X and Y are linearly related,
   $$X = aY + b$$
   for some constants (a, b) with $a \neq 0$.

   Example 3 If $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, then $\rho(X, Y) = \rho$.

   □

   As we have observed before, if X, Y are jointly Gaussian and $\rho = 0$, then they are independent. Otherwise, $\rho = 0$ does not imply independence.

Characteristic functions

The characteristic function is essentially the Fourier transform of the p.d.f. or p.m.f. They are useful in practice not for the usual reasons engineers use Fourier transforms (e.g., frequency content), but because they can provide a means of computing moments (as we will see), and they are useful in finding distributions of sums of independent random variables.

Definition 9 Let X be a r.v. The **characteristic function** (ch.f.) of X is

$$\boxed{\phi_X(u) = E[e^{iux}]}$$

for $u \in \mathbb{R}$. (Here, $i = \sqrt{-1}$. We will not use $\sqrt{-1} = j$.) □

Let us write some more explicit formulas. Suppose X is a continuous random variable. Then (by the law of the unconcious statistician)

$$\boxed{\phi_X(u) = \int_{-\infty}^{\infty} e^{iux} f_X(x)\, dx.}$$

This may be recognized as the Fourier transform of $f_X(x)$, where u is the "frequency" variable. (Comment on sign of exponent.) Note that given ϕ_X we can determine f_X by an inverse Fourier transform:

If X is a discrete r.v.,

$$\boxed{\phi_X(u) = \sum_i e^{iux_i} p_X(x_i),}$$

which we recognize as the discrete-time Fourier transform, and as before u is the "frequency" variable. (Comment on the sign of the exponent.) Given a ϕ_X, we can find p_X by the inverse discrete-time Fourier transform.

Properties:

1. $\phi_X(0) = 1$. (Why?)
2. $|\phi_X(u)| \leq 1 \forall u$. (Why?)
3. ϕ_X and f_X form a unique Fourier transform pair.
   $$f_X \leftrightarrow \phi_X.$$
   Thus, ϕ_X provides yet another way of displaying the probability structure of X.
4. $\phi_X(u) = \int_{-\infty}^{\infty} e^{iux} dF_X(x)$. This is referred to as the Fourier-Stieltjes transform of F_X.
5. ϕ_X is uniformly continuous.

Definition 10 For an r.v. X, the kth **moment** of X is $E[X^k]$, for $k \in \mathbb{N}$. □
We can write
$$E[X^k] = \int_{-\infty}^{\infty} x^k dF_X(x).$$

Theorem 1
If $E[|X|^k] < \infty$ *then*
$$\boxed{E[X^k] = i^{-k} \frac{d^k}{du^k} \phi_X(u) \Big|_{u=0}.}$$
That is, we can obtain moments by differentiating the characteristic function. For this reason, characteristic functions (or functions which are very similarly defined) are sometimes referred to as **moment generating functions**.

Proof
$$\phi_X(u) = E[e^{iux}] = E \sum_{k=0}^{\infty} \frac{(iu)^k}{k!} x^k = \sum_{k=0}^{\infty} \frac{(iu)^k}{k!} E[x^k].$$
Then
$$\frac{d^j}{du^j} \phi_X(u) = \frac{j! i^j}{j!} E[x^j] + \sum_{k>j} u^{k-j} (\text{other stuff})$$
so at $u = 0$,
$$\frac{d^j}{du^j} \phi_X(j) \Big|_{u=0} = i^j E[X^j].$$
□

Example 4 $X \sim \mathcal{N}(\mu, \sigma^2)$. Then it can be shown (homework!) that
$$\phi_X(u) = e^{-u^2\sigma^2/2 + iu\mu}$$
Then it is straightforward to verity (homework!) that
$$E[X] = \mu \qquad E[X^2] = \sigma^2 + \mu^2$$
so that $\text{var}(X) = \sigma^2$. □

Definition 11 For a joint r.v. (X, Y) we define a **joint characteristic function**
$$\phi_{XY}(u, v) = E[e^{iuX + ivY}].$$
□

Then ϕ_{XY} and F_{XY} are uniquely related (two-dimensional Fourier transforms).

Definition 12 The nth order moments of two random variables are the quantities of the form
$$\mu_{k,l} = E[X^k Y^l] \qquad k \geq 0, l \geq 0, k + l = n.$$
The **nth order central moments** are
$$m_{kl} = E[(X - E[X])^k (Y - E[Y])^l \qquad k \geq 0, l \geq 0, k + l = n.$$
□

Example 5 For $n = 2$, the second order moments are $E[X^2]$, $E[Y^2]$ and $E[XY]$.
The central moments are $\text{cov}(X, Y)$, $\text{var}(X)$ and $\text{var}(Y)$. □
Properties:

1. Moments:
$$\mu_{k,l} = \frac{\partial^n}{\partial u^k \partial v^l} \phi_{XY}(u, v) \Big|_{u,v=0} i^{-(k+l)}$$
2. X and Y are independent if and only if $\phi_{X,Y}(u, v) = \phi_X(u)\phi_Y(v)$ for all $(u, v) \in \mathbb{R}^2$.

Sums of independent random variables

Let X and Y be independent r.v.s, and let

$$Z = X + Y.$$

Then

$$\phi_Z(u) = E[\exp(iuz)] = E[\exp(iuX + iuY)] = \phi_{X,Y}(u,u).$$

But also

$$E[\exp(iuX + iuY)] = E[\exp(iuX)\exp(iuY)] = \phi_X(u)\phi_Y(u).$$

So

$$\boxed{\phi_Z(u) = \phi_X(u)\phi_Y(u)}$$

If X and Y are continuous r.v.s, then so is Z.

$$f_Z(z) = \mathcal{F}^{-1}[\phi_Z(u)] = \mathcal{F}^{-1}[\phi_X(z)\phi_Y(z)] = f_X(z) * f_Y(z)$$

by the **convolution theorem**.

Thus, **when continuous independent random variables are added, the p.d.f of the sum is the convolution of the p.d.f.s** (and respectively p.m.f. for discrete independent r.v.s).

An example: Jointly Gaussian

If $(X,Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, then

$$\phi_{X,Y}(u,v) = \exp[i(u\mu_x + v\mu_y) - \frac{1}{2}(u^2\sigma_x^2 + v^2\sigma_y^2 + 2uv\rho\sigma_x\sigma_y)]$$

We make an observation here: the "form" of the Gaussian p.d.f. is the exponential of quadratics. The form of the Fourier transform of the eponential of quadratics is of the form exponential of quadratics. This little fact gives rise to much of the analytical and practical usefulness of Gaussian r.v.s.

Characteristic functions marginals

We observe that

$$\phi_{X,Y}(u,0) = \phi_X(u)$$

In our Gaussian example, we have

$$\phi_X(u) = \phi_{X,Y}(u,0) = \exp(iu\mu_x - \sigma_x^2 u^2/2)$$

which is the ch.f. for a Gaussian,

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2).$$

We could, of course, have obtained a similar result via integration, but this is **much** easier.

Some important inequalities

In general, when we observe an outcome of a random variable, we "expect" it to be near the mean (that is, near the expected value). Further, the farther the outcome is from the mean, the less likely we expect the outcome to be. There are some very useful probabilities which quantize these intuitive "expectations." These are the Markov inequality, and its consequences, the Chebyshev inequality and the Chernoff bound. We will introduce these here.

Let $B \in \mathcal{B}$ for a Borel set $\mathcal{B}$. Recall that

$$I_B(x) = \begin{cases} 1 & x \in B \\ 0 & x \notin B. \end{cases}$$

Let X be a random variable, and let $Y = I_B(X)$. This is a measurable function, so Y is another random variable.

$$E[Y] = P_X(B) = P(X \in B).$$

We will use this "expectation as probability" idea to get a bound.

Suppose g is a nonnegative, nondecreasing function. Suppose $b \in \mathbb{R}$ with $g(b) > 0$. Consider the function

$$h(x) = \frac{g(x)}{g(b)} \rightarrow \begin{cases} \geq 1 & x \geq b \\ \geq 0 & \forall x. \end{cases}$$

Observe that

$$h(x) \geq I_{[b,\infty)}(x)$$

for all x, since

$$I_{[b,\infty)}(x) = \begin{cases} 1 & x \geq b \\ 0 & x < b. \end{cases}$$

Note that $h(b) = I_{[b,\infty)}(b)$.

Now we have

$$E[h(X)] \geq E[I_{[b,\infty)}(X)] = P(X \geq b).$$

Also,

$$E[h(X)] = \frac{E[g(X)]}{g(b)}$$

Thus

$$P(X \geq b) \leq \frac{E[g(X)]}{g(b)}.$$

A similar result can be established if g is nonnegative, nondecreasing on $[0, \infty)$ and symmetric about 0. We can thus establish that

$$\boxed{P(|X| \geq b) \leq \frac{E[g(X)]}{g(b)}.}$$

Special case: Assume $X \geq 0$, and let

$$g(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0. \end{cases}$$

The inequality above gives rise to the **Markov inequality**:

$$\boxed{P(X \geq b) \leq \frac{E[X]}{b}}$$

for all $b > 0$. Somewhat more generally, the Markov inequality says

$$\boxed{P(|X| \geq b) \leq \frac{E[|X|]}{b}}$$

for any $b > 0$.

Special Case: Take $g(x) = X^2$. (This satisfies the requirements for g.) Then

$$P(|Y| > b) \leq \frac{E[Y^2]}{b^2}.$$

Let $Y = X - \mu_x$. We obtain the **Chebyshev inequality**,

$$\boxed{P(|X - \mu_x| > b) \leq \frac{E[(X - \mu_x)^2]}{b^2} = \frac{\mathrm{var}(X)}{b^2}.}$$

Interpretation: The probability that X differs from its mean by more than some amount b is less than the variance of X over b^2. Further away, less probable. Higher varianace, more probable.

Special case: The **Chernoff bound**. In this case, let us take X positive, and let $g(x) = e^{sx}$ for $s > 0$. Then we obtain

$$P(X \geq b) \leq \frac{E[e^{sx}]}{e^{sb}}.$$

There is some flexibility in the choice of s, which may be selected to make the bound as tight as possible.

The Chernoff bound is a powerful tool which has been put to good use in digital communications. (See, e.g., Wozencraft and Jacobs.)

Jensen's inequality

Jensen's inequality can be used in some cases to "interchange" expectation and function evaluation (at least, approximately). It is based on the idea of convex functions.

Definition 13 A function $c : \mathbb{R} \to \mathbb{R}$ is **convex** if

$$c((1 - \alpha)x + \alpha y) \leq (1 - \alpha)c(x) + \alpha c(y).$$

for all $x, y\ in \mathbb{R}$ and $0 \leq \alpha \leq 1$. □

That is, a function is convex if the chord connecting the points $(x, f(x))$ and $(y, f(y))$ lies *above* the function between x and y. (Draw a picture.) It can be shown that if c is twice differentiable then c is convex iff $c''(x) \geq 0 \forall x \in \mathbb{R}$.

Example 6 $c(x) = e^x$
$c(x) = e^{-x}$
$c(x) = x^2$.
$c(x) = ax + b$ □

It can be shown that all convex functions are measurable with respect to the Borel field $\mathcal{B}(\mathbb{R})$.

Theorem 2
If $c : \mathbb{R} \to \mathbb{R}$ is convex, then

$$\boxed{E[c(x)] \geq c(E[x]),}$$

with equality if and only iff $c(x)$ is a constant function.

In otherwords, we can interchange expectation and functions, and at least get a bound.

Example 7 $E[X^2] \geq (E[X])^2$
$E[e^X] \geq e^{E[X]}$
If $X \geq 0$, $E[1/X] \geq 1/E[X]$
If $X > 0$, $-E[\log X] \geq -\log E[x]$. □

Cauchy-Schwartz inequality

This is an inequality that holds in any Hilbert space. It more or less forms the theme for the first several weeks of 6030.

Theorem 3
If $E[X^2] < \infty$ *and* $E[Y^2] < \infty$ *then*

$$|E[XY]|^2 \leq E[X^2]E[Y^2].$$

For example,

$$|\operatorname{cov}(X,Y)|^2 \leq \operatorname{var}(X)\operatorname{var}(Y)$$

implying $|\rho| \leq 1$.

Observe that $\operatorname{var}(X) = E[X^2] - E[X]^2 \geq 0$ using Schwartz and Jensen inequalities.

Conditional Expectations and Distributions

Suppose X is a discrete r.v. Now we define the conditional distribution of another r.v. Y given $X = x_k$ (at some point where $P(X = x_k) > 0$) by

$$\boxed{F_{Y|X}(y|x_k) = P(Y \leq y|X = x_k) = \frac{P(Y \leq y, X = x_k)}{P(X = x_k)}.}$$

By the law of total probability,

$$F_Y(y) = \sum_k F_{Y|X}(y|x_k)p_X(x_k).$$

As we discussed before, when we condition on an event, we are shrinking the sample space under consideration. So there is some normalization that takes place.

We also define

$$E[Y|X = x_k] = \int_{-\infty}^{\infty} y dF_{Y|X}(y|x_k).$$

Note that this depends on the value of x_k; it is a function of x_k. Let us now take the expectation with respect to X:

$$E_X[E[Y|X = x_k]] = \sum_k E[Y|X = x_k]P_X(x_k) = E[Y].$$

We can think of $E[Y|X = x_k]$ as a discrete random variable that is a function of X.

For a discrete r.v. X, the function $F_{Y|X}(y|x_k)$ could be either a discrete or a continuous r.v. Discrete:

$$p_{Y|X}(y|x_k) = P(Y = y|X = x_k)$$

Continuous: There exists a function $f_{Y|X}$ such that

$$F_{Y|X}(y|x_k) = \int_{-\infty}^{y} f_{Y|X}(z|x_k)\,dz.$$

We can also write

$$F_{Y|X}(y|x_k) = E[I_{(-\infty,y]}(Y)|X = x_k]$$

If Y is discrete we have

$$p_{Y|X}(y|x_k) = \frac{P(Y = y, X = x_k)}{p_x(x_k)} = \frac{p_{xy}(x_k, y)}{p_x(x_k)}.$$

When X is a continuous r.v., conditional probabilities and expectations are somewhat more complicated, because $P(X = xk) = 0$ for any particular value of x.

Recall that $E[Y|X_k] = g(x_k)$ for some function g, and $E[g(x)] = E[Y]$.

Definition 14 Suppose Y is an r.v. on the probability space $(\Omega, \mathcal{F}, P)$ with $E[|Y|] < \infty$. Then for $A \in \mathcal{F}$, define

$$\int_A Y dP = E[I_A(Y)].$$

$\square$

Definition 15 Suppose X and Y are random variables and $E[|Y|] < \infty$. The conditional expectation of Y given $X = x$ is any measurable function $g(x) = E[Y|X = x]$ of x satisfying

$$\int_B E[Y|X = x] P_X(dx) = \int_{X^{-1}(B)} Y dP \tag{2}$$

for all $B \in \mathcal{B}$, where $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. $\square$

1. It can be shown that under the stated conditions, such a function always exists.
2. If X is discrete then $E[Y|X = x_k]$ as defined earlier satisfies the property.
3. $E[Y|X = x]$ is unique, in the sense that if there are two functions $g(x)$ and $h(x)$ both satisfying (2) then $P(g(x) = h(x)) = 1$.

When a condition is true with probability 1, we say that it is true "almost surely," or "a.s."

Once we have defined conditional expectation, we can define a conditional c.d.f.:

$$F_{Y|X}(y|x) = E[I_{(\infty,y]}(y)|X = x].$$

Properties:

1. This definition agrees with the previous one when X is discrete.
2. $F_Y(y) = \int_{\mathbb{R}} F_{Y|X}(y|x) P_X(dx)$.
3. $F_{Y|X}$ is a c.d.f. as a function of y because it satisfies all the properties of a c.d.f.
4. If X and Y are jointly continuous then $F_{Y|X}(y|x)$ has a density for every x,

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

There is another interpretation:

$$\begin{aligned}
F_{Y|X}(y|x) &= \lim_{\Delta x \to 0^+} P(Y \leq y | x - \Delta x/2 < X \leq x + \Delta x/2) \\
&= \frac{\lim_{\Delta x \to 0^+} P(Y \leq y, x - \Delta/2 < X \leq x + \Delta x/2)/\Delta x}{\lim_{\Delta x \to 0^+} P(x - \Delta/2 < X \leq x + \Delta x/2)/\Delta x} \\
&= \frac{\frac{\partial}{\partial x} F_{XY}(x,y)}{\frac{\partial}{\partial x} F_X(x)} \\
&= \frac{\frac{\partial}{\partial x} F_{XY}(x,y)}{f_x(x)}
\end{aligned}$$

If X and Y are jointly continuous then $\frac{\partial}{\partial y} F_{Y|X}(y|x)$ exists and

$$\boxed{\frac{\partial}{\partial y} F_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}.}$$

Also,

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy.$$

Analogously for continuous random variables

$$E[Y|X = x_k] = \sum_l y_l p_{Y|X}(y_l|x_k).$$

A more general definition of conditional expectation

We will explore conditional expectation in terms of probability spaces. Suppos $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$. A function $g : \Omega \to \mathbb{R}$ is "measurable with respect to $\mathcal{G}$" if

$$\{\omega \in \Omega : g(\omega) \in B\} \in \mathcal{G} \text{ for all } B \in \mathcal{B}$$

(Any such function $g()$ would be a r.v. too. But this $g()$ is more restricted.)

We will now define "conditioning on a σ-field." Suppose Y is a r.v. with $E[|Y|^2] < \infty$ and $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$. Then $E[Y|\mathcal{G}]$ is any $\mathcal{G}$-measurable random variable such that

$$\int_A E[Y|\mathcal{G}] dP = \int_A Y dP \text{ for all } A \in \mathcal{G}.$$

If Y itself were $\mathcal{G}$ measurable it would be its own conditional expectation.

Example 8 Let $(\omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. Let $\mathcal{G} = \{(-\infty, 0), [0, \infty), \emptyset, \mathbb{R}\}$. (This is a σ-field.) Let Y be a r.v. Then $E[Y|\mathcal{G}]$ is any $\mathcal{G}$-measurable r.v. satisfying

$$\int_{\mathbb{R}} E[Y|\mathcal{G}] dP = E[Y]$$

$$\int_{(-\infty,0)} E[Y|\mathcal{G}] dP = \int_{(-\infty,0)} Y dP$$

$$\int_{[0,\infty)} E[Y|\mathcal{G}] dP = \int_{[0,\infty)} Y dP.$$

Note: to be measurable on $\mathcal{G}$ means being constant on $(-\infty, 0)$ and $[0, \infty)$.

The $\mathcal{G}$-measurable r.v.s in this case are simple functions:

$$g(\omega) = \begin{cases} b_+ & \omega \geq 0 \\ b_- & \omega < 0 \end{cases}$$

so

$$b_- P((-\infty, 0)) = \int_{(-\infty,0)} Y dP$$

$$b_+ P([0, \infty)) = \int_{[0,\infty)} Y dP$$

This gives us two equations in two unknowns.

$$b_+ = \frac{\int_{[0,\infty)} Y dP}{\int_{[0,\infty)} dP} \qquad b_- = \frac{\int_{-\infty,0)} Y dP}{\int_{(-\infty,))} dP}$$

$\square$

Definition 16 If X is an r.v., define $\sigma(X)$ (the σ-field generated by X) to be

$$\{\{\omega \in \Omega : X(\omega) \in B\} \text{ for } B \in \mathcal{B}\}.$$

$\square$

Fact: A r.v. Y is measurable with respect to $\sigma(X)$ if and only if there is a measurable function $g : \mathbb{R} \to \mathbb{R}$ such that $Y = g(X)$.

We now define conditional expectation with respect to a σ-field:

Definition 17 If X and Y are r.v.s with $E[|Y|] < \infty$, we define

$$E[Y|X] = E[Y|\sigma(X)]$$

$\square$

Properties:

1. By the fact stated above, we can write

   $$E[Y|X] = g(x)$$

   for some function g, $g(x) = E[Y|X = x]$.

2. $E[Y] = E[E[Y|X]]$
3. If Y itself is G-measurable, then $E[Y|\mathcal{G}] = Y$.
4. $E[\alpha Y_1 + \beta Y_2|\mathcal{G}] = \alpha E[Y_1|\mathcal{G}] + \beta E[Y_2|\mathcal{G}]$.
5. If $Y \geq 0$, then $E[Y|\mathcal{G}] \geq 0$.
6. If $E[|Y|] < \infty$ and $\mathcal{G} \subset \mathcal{E} \subset \mathcal{F}$, then

   $$E[E[Y|\mathcal{E}]] = E[Y|\mathcal{G}].$$

   Idea: If you first condition on a field that is less “course” than $\mathcal{G}$ you get a r.v. Then condition on $\mathcal{G}$.

Definition 18 Two σ-fields $\mathcal{G}$ and $\mathcal{H}$ are **independent** if

$$P(GH) = P(G)P(H) \text{ for all } G \in \mathcal{G} \text{ and } H \in \mathcal{H}.$$

$\square$

Note: X and Y are independent r.v.s iff $\sigma(X)$ and $\sigma(Y)$ are independent σ-fields.

6. If $\sigma(Y)$ is independent of $\mathcal{G}$ then $E[Y|\mathcal{G}] = E[Y]$.
7. If Y is $\mathcal{G}$-measurable then $E[Y|\mathcal{G}] = E[Y]$.

   So, for example, if $\mathcal{G} = \sigma(X)$, and $Y = g(X)$ for some $g : \mathbb{R} \to \mathbb{R}$ (that is, Y is $\mathcal{G}$-measurable) then

   $$E[Y|X] = Y = g(X).$$

   More informally, $E[g(x)|X] = g(x)$.