2.3 Scalar summaries

Theorem 2.7. (Bivariate LOTUS)

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\,\mathrm{d}x\mathrm{d}y \tag{2.24}$$

marginal moments

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x \tag{2.25}$$

conditional moments

$$\mathbb{E}[g(Y) \mid X = x] = \int_{-\infty}^{\infty} g(y) f_{Y \mid X}(y \mid x)\,\mathrm{d}y \tag{2.26}$$

Recall that if X and Y are independent, we have $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. This means that if we multiply any two functions $g_1(X)$ and $g_2(Y)$, the expectation of their product is the product of their expectations. Indeed, using (2.24) with the function $g(x, y) = g_1(x)g_2(y)$ gives

$$\begin{aligned}
\mathbb{E}[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y) f_{X,Y}(x, y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y) f_X(x) f_Y(y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty} g_2(y) f_Y(y) \left\{ \int_{-\infty}^{\infty} g_1(x) f_X(x)\,\mathrm{d}x \right\} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} g_2(y) f_Y(y) \mathbb{E}[g_1(X)] \mathrm{d}y = \mathbb{E}[g_2(Y)]\mathbb{E}[g_1(X)].
\end{aligned}$$

2.4 Linear combinations of random variables

We've studied joint distributions for pairs of random variables X and Y, but there's nothing special about two. What about joint distributions for n random variables X_1, X_2, X_3, ..., X_n? We won't study these in full generality, and instead will focus on the special case where $X_1, \ldots, X_n$ are *independent*.

Definition 2.11. We say the random variables $X_1, \ldots, X_n$ are independent if for any functions $g_1, \ldots, g_n$ such that the below quantities are finite, we have

$$\mathbb{E}[g_1(X_1)g_2(X_2)\ldots g_n(X_n)] = \mathbb{E}[g_1(X_1)]\mathbb{E}[g_2(X_2)]\ldots\mathbb{E}[g_n(X_n)]. \tag{2.27}$$

A few remarks on this definition:

- Consider the case $n = 2$, and call the two random variables X and Y. The above definition states that X and Y are independent if for every g_1, g_2, we have $\mathbb{E}[g_1(X)g_2(Y)] =$

$\mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)]$. But before, we *defined* independence of two random variables as the statement that their joint pdf/pmf is the product of the marginals and *concluded* that $\mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)]$ using bivariate LOTUS. This is not a contradiction: one can show that

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \Longleftrightarrow \quad \mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)] \text{ for all } g_1, g_2.$$

- For $n > 2$, we have an analogous equivalence between (2.27) and the statement that the joint pdf/pmf of the n random variables factorizes as the product of the n marginals. However, we will not go into detail about joint pdfs/pmfs of $n > 2$ random variables.

Next, we consider an even more specialized case.

Definition 2.12. A collection of random variables X_1, X_2, X_3, ..., X_n are **independent and identically distributed** (**iid**) if they are independent, and if their marginal distributions are all identical. We denote this

$$X_1, X_2, ..., X_n \overset{\text{iid}}{\sim} P, \tag{2.28}$$

where P is generic notation for the identical marginal distribution that each X_i follows. Instead of P, we might sometimes use F, f, or M if context makes clear that the distribution is represented by a particular cdf, pdf, mgf, etc.

Remark 2.1. “Identically distributed” does not mean “identical.” Each X_i is a randomly determined quantity that will have its own distinct value, but the probability rules governing what that value will be are identical.

Given an iid collection of random variables X_i, we are interested in the behavior of their *linear combination*, which is itself a new random variable:

$$T_n = \sum_{i=1}^{n} c_i X_i.$$

In particular, we will focus on the behavior of the simple sum and the average of iid random variables:

$$S_n = \sum_{i=1}^{n} X_i$$
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}S_n.$$

Understanding the behavior of these two random quantities will be very important when we start studying statistics.

So, in what follows we will assume $X_1, X_2, ..., X_n \overset{\text{iid}}{\sim} P$ and introduce the following generic notation:

- $\mu = \mathbb{E}[X_1]$
- $\sigma^2 = \text{var}(X_1)$;

- $M(t) = \mathbb{E}[e^{tX}]$.

Because the X_i are iid, these objects are the same for each X_i, so μ, σ^2, and M denote the common mean, variance, and mgf that they all share. We are implicitly assuming that these objects exist to begin with, but recall that this is not a forgone conclusion. The Cauchy distribution, for instance, has none of these objects, but it remains a perfectly coherent and useful distribution. Anyway, given this setup, we want to know the properties of S_n and $\bar{X}_n$. In particular...

- What are the means of S_n and $\bar{X}_n$?
- What are the variances of S_n and $\bar{X}_n$?
- What are the full distributions of S_n and $\bar{X}_n$?
- What happens to S_n and $\bar{X}_n$ when $n \to \infty$?

We'll cover the first three questions in this lecture, and the fourth question in the next lecture.

2.4.1 What's the mean of the sum and the average?

Theorem 2.8 (Linearity of Expectation)**.** Suppose...

- $X_1, X_2, \dots, X_n$ are possibly dependent random variables with finite means: $E[|X_i|] < \infty$ for all $i = 1, \dots, n$;
- $a_1, a_2, \dots, a_n \in \mathbb{R}$ are constants;
- X_i could have any type, any distribution, any dependence.

Then

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i].$$

Partial proof. Consider the $n = 2$ case with jointly absolutely continuous $(X, Y) \sim f_{XY}$, where f_{XY} is the joint pdf, and X, Y could be dependent. For any $a, b \in \mathbb{R}$, we want to show that $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. Recall multivariable LOTUS: $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dx\, dy$. We apply this to the linear transformation $g(x, y) = ax + by$:

$$\begin{aligned}
\mathbb{E}[aX + bY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by)\, f_{XY}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ax\, f_{XY}(x, y) + by\, f_{XY}(x, y)\right] dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ax\, f_{XY}(x, y)\, dx\, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} by\, f_{XY}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ax\, f_{XY}(x, y)\, dy\, dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} by\, f_{XY}(x, y)\, dx\, dy \\
&= a \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx + b \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy \\
&= a \int_{-\infty}^{\infty} x\, f_X(x)\, dx + b \int_{-\infty}^{\infty} y\, f_Y(y)\, dy \\
&= a\, \mathbb{E}[X] + b\, \mathbb{E}[Y].
\end{aligned}$$

This covers the $n = 2$ case. The general case follows by induction. $\square$

Corollaries.

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mu = \boxed{n\mu.}$$

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\tfrac{1}{n} S_n\right] = \frac{1}{n}\mathbb{E}[S_n] = \boxed{\mu.}$$

So, when $X_1, X_2, \dots, X_n \overset{\text{iid}}{\sim} P$ with common mean $\mu = \mathbb{E}[X_1]$, then no matter what P is, we know

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu, \qquad \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu.$$

2.4.2 What's the variance?

Theorem 2.9. If $a, b \in \mathbb{R}$ constant and X, Y are independent random variables, then

$$\operatorname{var}(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y).$$

Proof. Shifting X and Y by any constant values won't change the variance of $aX + bY$, X, or Y. This means that to prove the theorem, it suffices to prove it in the case that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. In this case, we know $\mathbb{E}[aX + bY] = 0$ as well, by linearity of expectation. Then $\operatorname{var}(aX + bY) =$

$\mathbb{E}[(aX + bY)^2]$, $\text{var}(X) = \mathbb{E}[X^2]$, and $\text{var}(Y) = \mathbb{E}[Y^2]$.

$$\begin{aligned}\text{var}(aX + bY) &= \mathbb{E}\left[\left(aX + bY\right)^2\right] \\ &= \mathbb{E}\left[a^2X^2 + b^2Y^2 + 2abXY\right] \\ &= a^2\mathbb{E}\left[X^2\right] + b^2\mathbb{E}\left[Y^2\right] + 2ab\,\mathbb{E}[XY] \\ &= a^2\,\text{var}(X) + b^2\,\text{var}(Y) + 2ab\mathbb{E}[X]\mathbb{E}[Y] \\ &= a^2\,\text{var}(X) + b^2\,\text{var}(Y).\end{aligned}$$

To get the second-to-last line, we used independence of X and Y, and the fact that $\text{var}(X) = \mathbb{E}[X^2]$ and $\text{var}(Y) = \mathbb{E}[Y^2]$ (recalling $\mathbb{E}[X] = \mathbb{E}[Y] = 0$). To get the last line, we again used that the first moments of X and Y are zero. □

Corollary 2.10. If $X_1, X_2, \dots, X_n$ independent, then by induction,

$$\text{var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2\,\text{var}(X_i).$$

We now use the corollary to compute the variance of the sum and average of n iid random variables $X_1, \dots, X_n$. Recall that $\text{var}(X_1) = \sigma^2$, so

$$\text{var}(S_n) = \text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) = \sum_{i=1}^n \sigma^2 = n\sigma^2,$$

and

$$\text{var}(\bar{X}_n) = \text{var}\left(\tfrac{1}{n}S_n\right) = \left(\frac{1}{n}\right)^2 \text{var}(S_n) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

This is true for any P, as long as the X_i are independent and the variance is actually finite!

2.4.3 What is the entire distribution of S_n or $\bar{X}_n$?

For the mean and the variance of S_n and $\bar{X}_n$, we got generic results that applied no matter what the underlying P was. We won't quite be able to do that for this third question. We'll just work some examples. The following will be our main tool:

Recall: If $X \sim M_X$ (mgf), $a, b \in \mathbb{R}$ constant, and $Y = aX + b$, then

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

Theorem 2.11. If $X \sim M_X$ and $Y \sim M_Y$ are independent and $a, b \in \mathbb{R}$ are constant, then

$$M_{aX+bY}(t) = M_X(at)\,M_Y(bt).$$

Proof.

$$\begin{aligned}M_{aX+bY}(t) &= \mathbb{E}[e^{t(aX+bY)}] = \mathbb{E}[e^{atX+btY}] = \mathbb{E}[e^{atX}e^{btY}] \\ &= \mathbb{E}[e^{atX}]\,\mathbb{E}[e^{btY}] \quad \text{(by independence)} \\ &= M_X(at)\,M_Y(bt).\end{aligned}$$

□

Corollary 2.12. Let $X_1, X_2, \ldots, X_n$ be independent and $a_1, a_2, \ldots, a_n$ be constants. Then by induction,

$$M_{\sum_{i=1}^n a_i X_i}(t) = M_{X_1}(a_1 t)\, M_{X_2}(a_2 t) \cdots M_{X_n}(a_n t).$$

Finally, we suppose $X_1, \ldots, X_n$ are not only independent but also identically distributed, with $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} M$ (common mgf shared by each X_i). Then Corollary 2.12 gives

$$M_{S_n}(t) = \prod_{i=1}^n M(t) = M(t)^n$$

$$M_{\bar{X}_n}(t) = M_{\frac{1}{n}S_n}(t) = M_{S_n}\left(\tfrac{t}{n}\right) = M\left(\tfrac{t}{n}\right)^n$$

Based on these, we can determine the distribution of the sum and the average in special cases.

Example 2.8 (Poisson)**.** $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$, $\quad M(t) = e^{\lambda(e^t-1)}$:

$$M_{S_n}(t) = M(t)^n = \left[e^{\lambda(e^t-1)}\right]^n = e^{n\lambda(e^t-1)} \implies \quad S_n \sim \text{Poisson}(n\lambda).$$

Example 2.9 (Gamma)**.** $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$, $\quad M(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$:

$$M_{S_n}(t) = M(t)^n = \left[\left(\frac{\beta}{\beta-t}\right)^\alpha\right]^n = \left(\frac{\beta}{\beta-t}\right)^{n\alpha},$$
$$M_{\bar{X}_n}(t) = M(t/n)^n = \left[\left(\frac{\beta}{\beta-t/n}\right)^\alpha\right]^n = \left(\frac{n\beta}{n\beta-t}\right)^{n\alpha}.$$

So

$$S_n \sim \text{Gamma}(n\alpha,\ \beta), \qquad \bar{X}_n \sim \text{Gamma}(n\alpha,\ n\beta).$$

Example 2.10 (Normal)**.** $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\quad M(t) = e^{\mu t + \frac{\sigma^2}{2}t^2}$:

$$M_{S_n}(t) = M(t)^n = \left[e^{\mu t + \sigma^2 t^2/2}\right]^n = e^{n\mu t + n\sigma^2 t^2/2},$$
$$M_{\bar{X}_n}(t) = M(t/n)^n = \left[e^{\mu\frac{t}{n} + \frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2}\right]^n = e^{\mu t + \frac{\sigma^2}{n}\frac{t^2}{2}}$$

So

$$S_n \sim \mathcal{N}(n\mu,\ n\sigma^2), \qquad \bar{X}_n \sim \mathcal{N}\left(\mu,\ \sigma^2/n\right).$$

Example 2.11 (Bernoulli)**.** Suppose $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$. The mgf is given by

$$M(t) = E[e^{tX_1}] = (1-p)\, e^{t\cdot 0} + p\, e^{t\cdot 1} = (1-p) + p\, e^t.$$

So

$$M_{S_n}(t) = M(t)^n = \left[(1-p) + p\, e^t\right]^n.$$

Let's show that this is precisely the mgf of a Binom(n, p). So let $Y \sim$ Binom(n, p). Then

$$\begin{aligned}
M_Y(t) = E[e^{tY}] &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1-p)^{n-k} \\
&= \left[(1-p) + p\,e^t\right]^n \quad \text{(binomial theorem: } (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k)
\end{aligned}$$

We conclude that

$$S_n \sim \text{Binom}(n, p).$$

That's our first actual proof that the sum of iid Bernoullis is binomial.

2.5 Summary

- Start with $X_1, X_2, \dots, X_n \overset{\text{iid}}{\sim} P$ with common mean and variance $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \text{var}(X_1)$.
- Define new random variables:

$$S_n = \sum_{i=1}^{n} X_i, \qquad \bar{X}_n = \frac{1}{n} S_n.$$

- Means:

$$\mathbb{E}[S_n] = n\mu, \qquad \mathbb{E}[\bar{X}_n] = \mu.$$

- Variances:

$$\text{var}(S_n) = n\sigma^2, \qquad \text{var}(\bar{X}_n) = \sigma^2/n.$$

- Distribution:

$$\begin{aligned}
P = \text{Poisson}(\lambda) &\implies S_n \sim \text{Poisson}(n\lambda) \\
P = \text{Bernoulli}(p) &\implies S_n \sim \text{Binom}(n, p) \\
P = \text{Gamma}(\alpha, \beta) &\implies S_n \sim \text{Gamma}(n\alpha, \beta), \quad \bar{X}_n \sim \text{Gamma}(n\alpha, n\beta) \\
P = \mathcal{N}(\mu, \sigma^2) &\implies S_n \sim \mathcal{N}(n\mu, n\sigma^2), \quad \bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)
\end{aligned}$$