10: Linear Combinations of Random Variables

PSTAT 120A: Summer 2022

Ethan P. Marzban

July 18, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions
- Transformations of Random Variables
- Double Integrals
- Random Vectors and the basics of multivariate probability
- Independence of random variables, and covariance/correlation

Some more on Indicators

Theorem

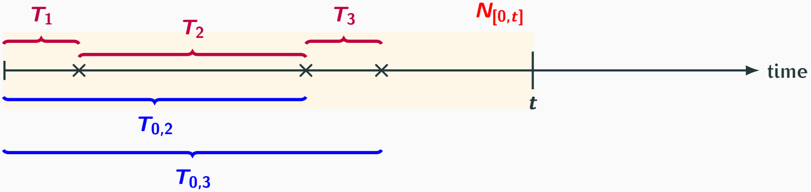
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two events $A, B \in \mathcal{F}$, then $\mathbb{1}_A \cdot \mathbb{1}_B = \mathbb{1}_{A \cap B}$

- Here's the general idea: both $\mathbb{1}_A$ and $\mathbb{1}_B$ are always either 0 or 1. Therefore, their product will also be 0 or 1.
- The product $\mathbb{1}_A \cdot \mathbb{1}_B$ will be 1 only when both $\mathbb{1}_A$ and $\mathbb{1}_B$ are 1; i.e. when both A and B occur. Otherwise, at least one of $\mathbb{1}_A$ and $\mathbb{1}_B$ will be zero. Hence, $\mathbb{1}_A \cdot \mathbb{1}_B = \mathbb{1}_{A \cap B}$.

Sums of Random Variables

Leadup: Poisson Processes

- Let's return to our Poisson Process for a moment.



- Recall that T_i denotes the time between the $(i-1)^{\text{th}}$ and i^{th} arrivals, and follows the $\text{Exp}(\lambda)$ distribution.
- Additionally, $T_{0,2}$ represents the time until the 2^{nd} arrival, and follows the $\text{Gamma}(2, \lambda)$ distribution.
- Now, if T_1 denotes the time until the first arrival, T_2 denotes the time between the first and second arrivals, and $T_{0,2}$ denotes the time until the second arrival, then it would appear

$$T_{0,2} = T_1 + T_2$$

- So, in some way, the "sum of two independent exponential distributions is distributed as Gamma…" This leads us into our next topic: sums of random variables.

Leadup

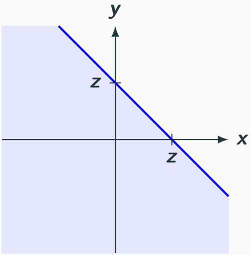
- Suppose we have two random variables X and Y that have a joint p.d.f. given by $f_{X,Y}(x, y)$.
- Now, consider the random variable $Z := X + Y$.
- First, note that Z is in fact a random variable. That is because both X and Y map from Ω to $\mathbb{R}$, meaning their sum will also map from $\Omega to \mathbb{R}$.
- Now, we know the mean and variance of Z:

$$\mathbb{E}[Z] = \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$
$$\text{Var}(Z) = \ = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

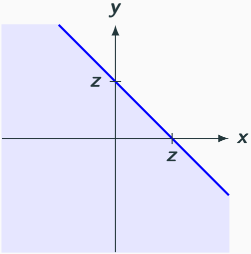
- Though that is all well and good, what if we are interested in the p.d.f. of Z?

Leadup

- So far, we really only have one tool to help us find the p.d.f. of a transformation of random variable(s): the CDF method!
- So, let's look at the c.d.f. of Z: $F_Z(z) = \mathbb{P}(X + Y \leq z)$.
- Very conveniently, we have already dealt with quantities like these! Specifically, we can compute them by computing a double integral of the joint $f_{X,Y}(x, y)$ over the region $\mathcal{R} = \{(x, y) : x + y \leq z\}$ for a fixed z.



Leadup



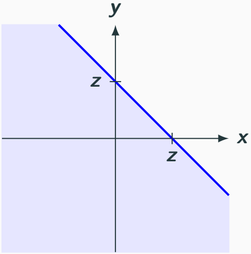
- So, using dy dx we see

$$F_Z(z) = \mathbb{P}(X + Y \leq z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y) \, \mathrm{d}y \, \mathrm{d}x$$

- Differentiating w.r.t z, and utilizing the Fundamental Theorem of Calculus, we find

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y) \, \mathrm{d}y \, \mathrm{d}x \right]$$
$$= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-x} f_{X,Y}(x, y) \, \mathrm{d}y \right) \, \mathrm{d}x$$
$$= \int_{-\infty}^{\infty} f_{X,Y}(x, z - x) \, \mathrm{d}x$$

Leadup



- If we had instead used dx dy we would have found

$$F_Z(z) = \mathbb{P}(X + Y \leq z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{X,Y}(x, y)\ \mathrm{d}x\ \mathrm{d}y$$

- Differentiating w.r.t z, and utilizing the Fundamental Theorem of Calculus, we find

$$\begin{aligned} f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_Z(z) &= \frac{\mathrm{d}}{\mathrm{d}z}\left[\int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{X,Y}(x, y)\ \mathrm{d}x\ \mathrm{d}y\right] \\ &= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-y} f_{X,Y}(x, y)\ \mathrm{d}x\right)\ \mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\ \mathrm{d}y \end{aligned}$$

P.D.F. of a Sum

- Thus, putting our work together, we have proven the following result:

Theorem

Given a pair of bivariate random variables (X, Y) with joint p.d.f. $f_{X,Y}(x, y)$, the p.d.f. of $Z := X + Y$ is given by

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x, z - x) \, \mathrm{d}x && (1) \\
&= \int_{-\infty}^{\infty} f_{X,Y}(z - y, y) \, \mathrm{d}y && (2)
\end{aligned}
$$

- Of course, often times the joint p.d.f. will be 0 over a significant portion of our region of integration. As such, often times the limits of our integrals will actually involve z. (We'll do an example in a bit.)

P.D.F. of a Sum

- An interesting simplification arises when $X \perp Y$. If $X \perp Y$, then $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ and the previous theorem becomes

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) \, \mathrm{d}x$$
$$= \int_{-\infty}^{\infty} f_X(z - y) \cdot f_Y(y) \, \mathrm{d}y$$

- Those of you with a bit of a math background might recognize this as the **convolution** of f_X and f_Y! That is,

$$X \perp Y \implies f_{X+Y} = (f_X * f_Y)$$

  and for this reason we sometimes refer to the previous theorem as the **convolution** formula.
  - As an aside: the convolution operator appears frequently through mathematics, especially in the context of functional analysis. Those of you who have taken a Differential Equations class might also recognize the convolution from there as well (in the context of Laplace Transformations!)

Example

Suppose $X, Y \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$. Identify the distribution of $Z := X + Y$, taking care to list out any/all relevant parameter(s)!

Discrete Convolution

- We have a similar result for discrete random variables:

Theorem: Discrete Convolution

Given a pair of bivariate random variables (X, Y) with joint p.m.f. $p_{X,Y}(x, y)$, the p.m.f. of $Z := X + Y$ is given by

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_{X,Y}(x, z - x) \tag{3}$$

$$= \sum_{y=-\infty}^{\infty} p_{X,Y}(z - y, y) \tag{4}$$

Example

Suppose $X, Y \overset{\text{i.i.d.}}{\sim} \text{Geom}(p)$. Identify the distribution of $Z := X + Y$, taking care to list out any/all relevant parameter(s)!

Comments

- The hardest part of these convolution problems is often finding the limits of integration/summation.
- I advise finding these limits by either drawing a picture, or appealing to indicators (like we did in the examples we did before).
- Additionally, I encourage you to keep the *derivation* of the convolution formula in mind as I find that to be quite helpful in determining the limits of integration! Let's see what I mean by way of an example:

Example

Let $X, Y \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$, and define $Z := X + Y$. Find $f_Z(z)$, the p.d.f. of Z. (As an aside: the distribution of Z is a special case of what is known as the **triangular distribution**.)

Using Indicators to Compute Expectations

Leadup

- Remember how we originally computed the expected value of the Binomial distribution? We used the definition of expectation $\mathbb{E}[X] = \sum_k k p_X(k)$ and then evaluated the sum.
- It wasn't terrible... You might ask, though, "is there another way to compute the mean of a Binomial distribution?"
- The answer is...

Figure 1: Source: https://knowyourmeme.com/memes/there-is-another

Example

- Here's the general idea. Let $X \sim \text{Bin}(n, p)$. We know then that X measures the number of successes in n Bernoulli(p) trials.
- Let's define n indicators in the following way:

$$\mathbb{1}_i := \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ trial resulted in a success} \\ 0 & \text{otherwise} \end{cases}$$

  for $i = 1, \cdots, n$.
- Note that

$$X = \sum_{i=1}^{n} \mathbb{1}_i$$

- Don't believe me? Well, consider $n = 2$. Then $X := \sum_{i=1}^{2} \mathbb{1}_i \in \{0, 1, 2\}$, where:
  - X is 0 when both indicators are zero; i.e. when both trials 1 and 2 resulted in a failure, which occurs with probability $(1 - p)^2$;
  - X is 1 when exactly one of the indicators is zero; i.e. when either trial 1 or trial 2 but not both resulted in a success, which occurs with probability $2p(1 - p)$;
  - X is 2 when both indicators are 1; i.e. when both trials 1 and 2 resulted in a success, which occurs with probability p^2.

Example

- Now, this is useful because we know that expectations are linear. That is,

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

- In general, we have the following fact:

Theorem

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and event $A \in \mathcal{F}$ and an indicator $\mathbb{1}_A$, then $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.

- Therefore, in our Binomial Problem,

$$\mathbb{E}[\mathbb{1}_i] = \mathbb{P}(\text{the } i^{\text{th}} \text{ trial resulted in a success}) = p$$

- So,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_i] = \sum_{i=1}^{n} (p) = np$$

Example

- Let's take this a step further! I claim we can get the variance of the binomial distribution quite simply using our indicator representation of X! Observe that:

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} \mathbb{1}_i\right)$$

- Now, since our trials are independent (by assumption), the indicators are all independent as well. This allows us to effectively pass the variance through the sum:

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} \mathbb{1}_i\right) = \sum_{i=1}^{n} \text{Var}(\mathbb{1}_i)$$

Example

- Let's take a bit of a detour.
- Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and event $A \in \mathcal{F}$, and an indicator $\mathbb{1}_A$, note that

$$\mathbb{1}_A^2 = \mathbb{1}_A \cdot \mathbb{1}_A = \left(\begin{cases} 1 & \text{if } A \\ 0 & \text{otherwise} \end{cases} \right) \cdot \left(\begin{cases} 1 & \text{if } A \\ 0 & \text{otherwise} \end{cases} \right)$$

$$= \begin{cases} 1 & \text{if } A \text{ and } A \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } A \\ 0 & \text{otherwise} \end{cases} = \mathbb{1}_A$$

- Therefore, $\mathbb{1}_A^2 = \mathbb{1}_A$ and

$$\text{Var}(\mathbb{1}_A) = \mathbb{E}[\mathbb{1}_A^2] - (\mathbb{E}[\mathbb{1}_A])^2 = \mathbb{E}[\mathbb{1}_A] - (\mathbb{E}[\mathbb{1}_A])^2 = \mathbb{P}(A) - \mathbb{P}(A)^2 = \mathbb{P}(A) \cdot \mathbb{P}(A^\complement)$$

- Let's make this a theorem:

Theorem

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and event $A \in \mathcal{F}$ and an indicator $\mathbb{1}_A$, then $\text{Var}(\mathbb{1}_A) = \mathbb{P}(A) \cdot \mathbb{P}(A^\complement)$.

Example

- So, going back to our Binomial problem,

$$\text{Var}(\mathbb{1}_i) = \mathbb{P}(\text{success on } i^{\text{th}} \text{ trial}) \cdot \mathbb{P}(\text{failure on } i^{\text{th}} \text{ trial}) = p(1-p)$$

- Therefore,

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(\mathbb{1}_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

- Quite a bit slicker than what we did before, don't you think?

Runs

- Let's use our newfound knowledge of indicators to tackle an interesting (yet fairly involved) concept.
- Suppose we toss a $p-$coin n times. One possible outcome is:

$$H\ H\ H\ T\ T\ H\ T\ H\ T\ T$$

- Note that there are some interesting patterns present in this outcome: specifically, notice that there are certain "chunks" or "blocks" of consecutive heads and tails:

$$\boxed{H\ H\ H}\ \boxed{T\ T}\ \boxed{H}\ \boxed{T}\ \boxed{H}\ \boxed{T\ T}$$

- This leads us to the notion of **runs**. A **run** is a string of consecutive heads (or tails); note that a run could be of length 1.
- An interesting question then arises: what can we say about the number of runs in a sequence of n coin tosses?
  - As an example: in our sample outcome above, we observed 6 runs.

Runs

- Again, suppose we toss a $p-$coin n times. For notational convenience, it will be useful to set $q := 1 - p$ to be the probability of the coin landing "tails."
- Let X denote the number of runs in these n tosses.

- **Question 1:** What is $\mathbb{E}[X]$?
- The key is to express X as a sum of suitably defined indicator random variables.
- To see how we can do this, let's examine the notion of runs a bit more closely. Specifically, what does it mean to say that the j^{th} toss (where $j \in \{1, 2, \cdots, n\}$) is the start of a new run?
- Well, if the j^{th} toss marks the beginning of a new run, clearly the result of the j^{th} toss must be different than the toss right before it!
- To that end, we define

$$\mathbb{1}_j := \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ and } (j-1)^{\text{th}} \text{ tosses were different} \\ 0 & \text{otherwise} \end{cases}$$

- For convenience, here is the definition of $\mathbb{1}_j$ again:

$$\mathbb{1}_j := \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ and } (j-1)^{\text{th}} \text{ tosses were different} \\ 0 & \text{otherwise} \end{cases}$$

- Now, how can we express X as a sum of some of these $\mathbb{1}_j$'s?
- First, note that it is not correct to write $X = \sum_{j=1}^{n} \mathbb{1}_j$. This is largely due to how we define $\mathbb{1}_1$: specifically, what is the 0^{th} toss?
- Additionally, we know that X is guaranteed to be at least 1, since we are guaranteed at least one run in our n tosses (all heads, or all tails).
- Therefore, we can see that the appropriate relationship between X and the $\mathbb{1}_j$'s is

$$X = 1 + \sum_{j=2}^{n} \mathbb{1}_j \tag{5}$$

- Therefore, by linearity,

$$\mathbb{E}[X] = 1 + \sum_{j=2}^{n} \mathbb{E}[\mathbb{1}_j]$$

Runs

- What is $\mathbb{E}[\mathbb{1}_j]$? Recall that $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ for any event A. Therefore,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_j] &= \mathbb{P}(\text{the } j^{\text{th}} \text{ and } (j-1)^{\text{th}} \text{ tosses were different}) \\ &= \mathbb{P}((j-1)^{\text{th}} \text{ toss was heads, } j^{\text{th}} \text{ toss was tails}) \\ &\quad + \mathbb{P}((j-1)^{\text{th}} \text{ toss was tails, } j^{\text{th}} \text{ toss was heads}) \\ &= pq + qp = 2pq\end{aligned}$$

- Therefore,

$$\begin{aligned}\mathbb{E}[X] &= 1 + \sum_{j=2}^{n} \mathbb{E}[\mathbb{1}_j] \\ &= 1 + \sum_{j=2}^{n} 2pq = \boxed{1 + 2(n-1)pq}\end{aligned}$$

  - Where did the $(n-1)$ come from? Note that there are $(n-2+1) = (n-1)$ terms in the sum!

- As an interesting extension: if the coin were fair, then

$$\mathbb{E}[X] = 1 + 2(n-1)\cdot\frac{1}{2}\cdot\frac{1}{2} = 1 + \frac{n-1}{2} = \frac{n+1}{2}$$

Runs

- **Question 1.5:** What value of p maximizes the number of runs?
- For notational convenience, set

$$f(p) := \mathbb{E}[X] = 1 + 2(n-1)p(1-p)$$

- We now differentiate, and set equal to zero:

$$\frac{\mathrm{d}}{\mathrm{d}p} f(p) = \frac{\mathrm{d}}{\mathrm{d}p}\left[1 + 2(n-1)p(1-p)\right] = 1 - 2p$$

$$\implies 1 - 2\hat{p} = 0 \implies \boxed{\hat{p} = \frac{1}{2}}$$

Runs

- **Question 2:** What is the PMF of X?
- Quite difficult, in general!
  - On the homework, you will investigate some of the simpler aspects of this question.

- **Question 3:** What is $\text{Var}(X)$?
- Again, a bit challenging- and, again, you'll work on this for homework!