Statistics 110: Introduction to Probability — Section 6, Fall 2018

MGFs and Joint, Conditional, and Marginal Distributions

TFs: Justin Zhu, Credits to Timothy Kang

Poisson Process

The Poisson process gives a story that links the Exponential distribution with the Poisson distribution. A Poisson process with rate λ has the following properties:

(1) The number of arrivals that occur in an interval of length t is distributed $\text{Pois}(\lambda t)$.

(2) The number of arrivals that occur in disjoint intervals are independent of each other.

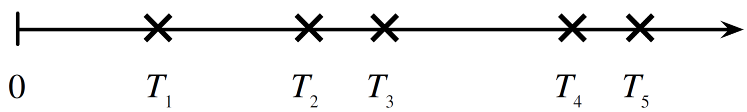

Figure 1: A Poisson Process. X marks the spot of an arrival. Source - Textbook

Count-Time Duality

Instead of asking how many events occur within some amount of time, we can flip the question around and ask how long it takes until some number of events occur. Let T_n be the amount of time it takes until the nth event occurs and let N_t be the number of events that occur within time t. What relationship do we have between T_n and N_t? And what is the intuition behind this

$$P(T_n > t) = P(N_t < n)$$

To reason about this in words, the event that the nth arrival time is greater than t is equivalent to the event that the number of arrivals by time t is less than n.

Using count-time duality, we can discern the distribution of T_n. Let us first look at the distribution for the first arrival.
For the first arrival, we have

$$P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - P(N_t < 1) = 1 - e^{-\lambda t}$$

This is the $\text{Expo}(\lambda)$ CDF!

Now that we know the distribution of T_1, what do we know about the distributions of the remaining inter-arrival times?

For $T_2 - T_1$, due to the independence of disjoint intervals, the past is irrelevant once the first arrival occurs, and hence $T_2 - T_1 \sim \text{Expo}(\lambda)$ by the same argument as before. Continuing in this manner, we can argue the same for subsequent inter-arrival times.

Can I Have a Moment?

Moment - Moments describe the shape of a distribution. The first three moments, are related to Mean, Variance, and Skewness of a distribution. The k^{th} moment of a random variable X is

$$\mu'_k = E(X^k)$$

What's a moment? Note that

Mean $\mu'_1 = E(X)$

Variance $\mu'_2 = E(X^2) = Var(X) + (\mu'_1)^2$

Mean, Variance, and other moments (Skewness, Kurtosis, etc.) can be expressed in terms of the moments of a random variable!

Moment Generating Functions

MGF For any random variable X, this expected value and function of dummy variable t;

$$M_X(t) = E(e^{tX})$$

is the **moment generating function (MGF)** of X if it exists for a finitely-sized interval centered around 0. Note that the MGF is just a function of a dummy variable t.

Why is it called the Moment Generating Function? Because the k^{th} derivative of the moment generating function evaluated 0 is the k^{th} moment of X!

$$\mu'_k = E(X^k) = M_X^{(k)}(0)$$

Why does this relationship hold?
By differentiation under the integral sign and then plugging in $t = 0$

$$M_X^{(k)}(t) = \frac{d^k}{dt^k} E(e^{tX}) = E(\frac{d^k}{dt^k} e^{tX}) = E(X^k e^{tX})$$
$$M_X^{(k)}(0) = E(X^k e^{0X}) = E(X^k) = \mu'_k$$

MGF of linear combination of X. If we have $Y = aX + c$, then

$$M_Y(t) = E(e^{t(aX+c)}) = e^{ct} E(e^{(at)X}) = e^{ct} M_X(at)$$

Uniqueness of the MGF. *If it exists, the MGF uniquely defines the distribution*. This means that for any two random variables X and Y, they are distributed the same (their CDFs/PDFs are equal) if and only if their MGF's are equal. You can't have different PDFs when you have two random variables that have the same MGF.

Summing Independent R.V.s by Multiplying MGFs. If X and Y are independent, then

$$M_{(X+Y)}(t) = E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) = M_X(t) \cdot M_Y(t)$$
$$M_{(X+Y)}(t) = M_X(t) \cdot M_Y(t)$$

The MGF of the sum of two random variables is the product of the MGFs of those two random variables.

Joint Distributions

Sometimes we have more than one random variable of interest, and we want to study probabilities associated with all of the random variables. Instead of studying the distributions of X_1, X_2, X_3 separately, we can study the distribution of the multivariate vector $\mathbf{X} = (X_1, X_2, X_3)$. Joint PDFs and CDFs are analogous to multivariate versions of univariate PDFs and CDFs. Usually joint PDFs and PMFs carry more information than the marginal ones do, because they account for the interactions between the various random variables. If, however, the random variables are independent, then the joint PMF/PDF is just the product of the marginals and we get no extra information by studying them jointly rather than marginally.

Joint Probability of events A and B: $P(A, B) = P(A \cap B)$

Joint CDF	**Joint PMF**	**Joint PDF**
$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$	$P(X = x, Y = y)$	$f_{X,Y}(x, y) = \frac{\delta}{\delta x}\frac{\delta}{\delta y}F_{X,Y}(x, y)$

Both the Joint PMF and Joint PDF must be non-negative and sum/integrate to 1. ($\sum_x \sum_y P(X = x, Y = y) = 1$) ($\int_x \int_y f_{X,Y}(x, y) = 1$). Like in the univariate case, you sum/integrate the PMF/PDF to get the CDF.

Conditional Distributions

By Bayes' Rule, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ Similar conditions apply to conditional distributions of random variables.

For discrete random variables:

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

For continuous random variables:

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

Marginal Distributions

Law of Total Probability says for an event A and partition $B_1, B_2, ...B_n$: $P(A) = \sum_i P(A \cap B_i)$, which implies that we can find the marginal probability by integrating out the B's.

To find the distribution of one (or more) random variables from a joint distribution, sum or integrate over the irrelevant random variables.

Getting the Marginal PMF from the Joint PMF

$$P(X = x) = \sum_y P(X = x, Y = y)$$

Getting the Marginal PDF from the Joint PMF

$$f_X(x) = \int_y f_{X,Y}(x, y)\, \mathrm{d}y$$

Independence of Random Variables

Review: A and B are independent if and only if either $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$.

Similar conditions apply to determine whether random variables are independent - two random variables are independent if their joint distribution function is simply the product of their marginal distributions, or that the a conditional distribution of is the same as its marginal distribution.

In words, random variables X and Y are independent for all x, y, if and only if one of the following hold:

- Joint PMF/PDF/CDFs are the product of the Marginal PMF
- Conditional distribution of X given Y is the same as the marginal distribution of X

$$
\begin{aligned}
X \text{ and } Y \text{ (discrete) are independent} &\iff P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \forall x, y \\
X \text{ and } Y \text{ (discrete) are independent} &\iff P(X = x, Y = y) = P(X = x)P(Y = y), \forall x, y \\
X \text{ and } Y \text{ (discrete) are independent} &\iff P(X = x|Y = y) = P(X = x), \forall x, y \\
X \text{ and } Y \text{ (continuous) are independent} &\iff P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \forall x, y \\
X \text{ and } Y \text{ (continuous) are independent} &\iff f_{X,Y}(x, y) = f_X(x)f_Y(y), \forall x, y \\
X \text{ and } Y \text{ (continuous) are independent} &\iff f_{X|Y}(X|Y) = f_X(X), \forall x, y
\end{aligned}
$$

Multivariate LotUS

In one dimension, we have: $E(g(X)) = \sum_x g(x)P(X = x)$, or $E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)\,\mathrm{d}x$

For discrete random variables:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)P(X = x, Y = y)$$

For continuous random variables:

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

Practice Problems

Example 1. Procrastinators.
Three students are working independently on their Stat 110 problem set. All three start at 1pm on the day the pset is due, and each takes an Exponential time with mean 6 hours to complete the homework. What is the earliest time when all three students will have completed the homework, on average?

Example 2. Counting Cars.
Cars pass by a certain point on a road according to a Poisson process with rate λ cars/minute. Let $N_t \sim \text{Pois}(\lambda t)$ be the number of cars that pass by that point in the time interval $[0, t]$, with t measured in minutes.

(a) A certain device is able to count cars as they pass by, but it does not record the arrival times. At time 0, the counter on the device is reset to 0. At time 3 minutes, the device is observed and it is found that exactly 1 car had passed by. Given this information, find the conditional CDF of when that car arrived. Also describe in words what the result says.

(b) In the late afternoon, you are counting blue cars. Each car that passes by is blue with probability b, independently of all other cars. Find the joint PMF and marginal PMFs of the number of blue cars and number of non-blue cars that pass by the point in 10 minutes.

Example 3. Linear Combination of Independent Normals.
Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Use the MGF to show that for any values of $a, b \neq 0$, $Y = aX_1 + bX_2$ is also Normal. (You may use the fact that the MGF of a $\mathcal{N}(\mu, \sigma^2)$ distribution is $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.)

Example 4. Return of Nick and Hibbert.
One of two doctors, Dr. Hibbert and Dr. Nick, is called upon to perform a series of n surgeries. Let H be the indicator random variable for Dr. Hibbert performing the surgeries, and suppose that $E(H) = p$. Given tht Dr. Hibbert is performing the surgeries, each surgery is successful with probability a, independently. Given that Dr. Nick is performing the surgeries, each surgery is successful with probability b independently. Let X be the number of successful surgeries.

(a) Find the joint PMF of H and X.

(b) Find the marginal PMF of X.

(c) Find the conditional PMF of H given $X = k$.

Example 5. Probabilities Using Joint CDFs and PDFs.
Let X and Y be continuous random variables. What is the probability that (X, Y) falls into the 2-D rectangle $[a_1, a_2] \times [b_1, b_2]$ in terms of a) the joint CDF of X and Y, $F(x, y)$, and b) the joint PDF of x and Y, $f(x, y)$?

Example 6. LogNormal Moments.
Let $Z \sim \mathcal{N}(0,1)$ and $Y = e^Z$. Then Y has a LogNormal distribution (because its log is Normal!).

(a) Find **all** the moments of Y. That is, find $E(Y^n)$ for all n.

(b) Show the MGF of Y does not exist, by arguing that the integral $E\big(e^{tY}\big)$ diverges for $t > 0$.