EE 126 — Probability and Random Processes

January 23, 2018 — Variance, Geometric Distribution

1 Lecture

Recap: Discrete Random Variables

- $P_X(k) = P(X = k)$
- $\mathbb{E}[X] = \sum_x xP(X = x) = \sum_{\omega \in \Omega} X(\omega)P(\omega)$
- $\mathbb{E}[g(X)] = \sum_x g(x)P(X = x)$
- $var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- A **joint PMF** takes the following form: $P(X = x, Y = y) = P(X = x \cap Y = y))$
- A **marginal PMF** takes the following form: $\sum_x P(X = x, Y = y) = P(Y = y)$
- $\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y)P(x = x, Y = y)$

Linearity of Expectation

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[aX + c] = a\mathbb{E}[X] + c$

This is a very powerful concept, especially when it comes to computing the expectation of a sum of random variables:

$$\mathbb{E}[X_1 + X_2 + ... + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + ... + \mathbb{E}[X_n]$$

Say we have $X \sim \text{Bin}(n,p)$, i.e. $X = X_1 + X_2 + ... + X_n$ where $X_i \sim \text{Bernoulli}(p)$, and we're looking for $\mathbb{E}[X]$. Linearity of expectation swiftly decrees that $\mathbb{E}[X] = np$.

Example: seats on a plane. It is a full flight. n passengers have assigned seats but ignore them and sit in random seats. What is $\mathbb{E}$[number of passengers who sit in their assigned seats]?

Let $X = \#$ passengers who sit in their assigned seats. Let $X_i = 1$ if passenger i sits in his own seat, and 0 otherwise. Then $X = X_1 + ... + X_n$, and

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X_1 + ... + X_n] \\
&= \mathbb{E}[X_1] + \mathbb{E}[X_2] + ... + \mathbb{E}[X_n] \\
&= n\mathbb{E}[X_1] \\
&= nP(X_1 = 1) \\
&= 1
\end{aligned}$$

This works even though the X_i's are not independent. In general, X_i's will not be independent, but the mean formula doesn't care!

Conditioning of Random Variables

Conditioning on random variables is similar to conditioning on events, where $P(X = k \mid A) = \frac{P(X=k \,\cap\, A)}{P(A)}$. In the random variable case, $P(X = k \mid Y = m) = \frac{P(X=k,\ Y=m)}{P(Y=m)} = P_{X|Y}(k \mid m)$. Note that $P_{X|Y}(k \mid m)$ is referred to as **the conditional PMF of X given Y.**

Incidentally, $X \mid Y$ is just another random variable! How do we know? Because $P(X \mid Y) \geq 0$, and $\sum_x P_{X|Y}(x \mid y) = \sum_x \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{1}{P_Y(y)} \sum_x P_{X,Y}(x,y) = \frac{1}{P_Y(y)} P_Y(y) = 1$.

Independence of Random Variables

Independence of random variables is similar to independence of events. If X and Y are independent random variables, $P_{X,Y}(x,y) = P_X(x)P_Y(y)\ \forall\ x, y$. This generalizes to more than two random variables.

If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. This often comes in handy!
Proof. $\mathbb{E}[XY] = \sum_x \sum_y xyP_{X,Y}(x,y) = \sum_x \sum_y xyP_X(x)P_Y(y) = (\sum_x xP_X(x))(\sum_y yP_Y(y)) = \mathbb{E}[X]\mathbb{E}[Y]$

Variance

As we've seen, the mean can be used as a single descriptor for a distribution. But it doesn't tell us the whole story. Another quantity we might ask for is the spread – maybe as the **variance**. The variance of a random variable X is defined as $var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Some "obvious" facts about the variance:

- $var(aX) = a^2 var(X)$
- $var(b) = 0$ (where b is a constant)
- $var(aX + b) = a^2 var(X)$

Also potentially important: $var(X - \mathbb{E}[X]) = var(X)$. The variance doesn't change if you strip the mean!

If X_1 and X_2 are independent random variables s.t. $X = X_1 + X_2$, then $var(X) = var(X_1) + var(X_2)$. If X and Y are dependent, $var(X) = var(X_1) + var(X_2) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])$ (where $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ is defined as the **covariance** between X and Y). We don't care about independence or dependence for the mean case, but we do care for the variance case!

Example: variance of a Bernoulli random variable. $X \sim \text{Bin}(n,p)$. What is its variance? We know that $X = X_1 + X_2 + ... + X_n$, where $X_i \sim \text{Bernoulli}(p)$ i.i.d. Since the random variables are independent, we have $var(X) = n \cdot var(X_1)$. So what is the variance of X_1? ($X_1 = 1$ with probability p, and 0 with probability $(1-p)$.) It follows that $var(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = p - p^2 = p(1-p)$, and therefore $var(X) = n \cdot var(X_1) = np(1-p)$.

Geometric Distribution

Example: the St. Petersburg Paradox. Let's play a game. I will pay you 2^k, where k is the number of flips of a fair coin it takes to get a heads. Thus $Y = 2^X$ where $X \sim \text{Geometric}(\frac{1}{2})$. For example, $TTTH \implies Y = 2^4 = 16$, and the payout of $TTTH$ would be \$16.

Paradoxically, the *expected* payout is infinite!

$$E[Y] = \sum_{k=1}^{\infty} 2^k P(X = k) = \sum_{k=1}^{\infty} 2^k 2^{-k} = \sum_{k=1}^{\infty} 1 = \infty$$

In general, $X \sim \text{Geom}(p)$ is a random variable. It is the "time to first success" random variable, so it often tells a nice story – e.g. first time to flip a heads, when flipping a coin independently with $P(H) = p$ on each flip (i.i.d. trials). In this example, $TTTH \implies X = 4$.

What is the PMF? (Visualize the story!) We find that the PMF is described by

$$\begin{aligned}
P(X = k) &= (1-p)^{k-1}p, && k = 1, 2, ... \\
P(X > k) &= (1-p)^k, && k = 1, 2, ... \\
P(X \leq k) &= 1 - (1-p)^k, && k = 1, 2, ...
\end{aligned}$$

Note: $P(X \leq k)$ is the CDF (the cumulative distribution function).

What is $\mathbb{E}[X]$? We will explore three methods of deriving the expectation (just because Prof. Ramchandran can).

Deriving $\mathbb{E}[X]$ (Method I: Naive)

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= p\sum_{k=1}^{\infty} k(1-p)^{k-1}
\end{aligned}$$

Note that $\sum_{k=0}^{\infty} \alpha^k = 1/(1-\alpha)$ if $|\alpha| < 1$.

Let $f(p) = \sum_{k=1}^{\infty}(1-p)^k = \sum_{k=0}^{\infty}(1-p)^k - 1 = \sum_{k=0}^{\infty} \frac{1-p}{p}$. Then $f'(p) = -\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{-p-(1-p)}{p^2} = -\frac{1}{p^2}$.

$$\mathbb{E}[X] = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p}$$

Deriving $\mathbb{E}[X]$ (Method II: Tail Sum Formula)

The tail sum formula (TSF) is defined as follows: if $X \geq 0$, $\mathbb{E}[X] = \sum_{k=1}^{\infty} P(X \geq k)$.

We thus have $\mathbb{E}[X] = \sum_{k=1}^{\infty}(1-p)^{k-1} = \frac{1}{p}$.

Proof of the TSF:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} kP(X = k) \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{k} 1 \cdot P(X = k) \\
&= \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} P(X = k) \\
&= \sum_{j=1}^{\infty} P(X \geq j)
\end{aligned}$$

Note: replacing k with a sum of k ones is an example of **lifting**. Lifting is "when you go to a higher level and you sort of see everything." Also, to quote Prof. Ramchandran, "when you have double summations, the first thing you should do is switch the order."

Deriving $\mathbb{E}[X]$ (Method III: Conditioning and "Renewal")

Here, we exploit the memorylessness property of the geometric distribution Geom(p).

Theorem (the memorylessness property). $P(X > n+m \mid X > n) = P(X > m),\ \ n, m > 0.$
Motivating question: if we've already waited n trials without success, what is the chance we'll wait m more trials without success? The idea is that the geometric distribution doesn't care how much we've failed so far – wherever we are, we've learned nothing since we started.

Proof.

$$\text{LHS} = \frac{P(X > n+m, X > n)}{P(X > n)} = \frac{P(X > n+m)}{P(X > n)} = \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = P(X > m) = \text{RHS}$$

Thus:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X|X=1]P(X=1) + \mathbb{E}[X|X>1]P(X>1) \\
&= \mathbb{E}[X|X=1]P(X=1) + (1+\mathbb{E}[X])P(X>1) \quad \text{(due to memoryless property)} \\
&= 1 \cdot p + (1+\mathbb{E}[X])(1-p)
\end{aligned}$$

Solving, we have

$$\begin{aligned}
\mathbb{E}[X] &= p + (1-p)(\mathbb{E}[X]+1) \\
0 &= -p\mathbb{E}[X] + 1 \\
\mathbb{E}[X] &= \frac{1}{p}
\end{aligned}$$

Coupon Collection

Example: the coupon collector problem. We need to get one of each coupon from a set of n coupons. We have to buy cereal boxes to get these coupons. What is the expected number of cereal boxes we'll have to purchase in order to collect all n distinct coupons? (In other words, what is $\mathbb{E}[\#$ cereal boxes needed to buy n distinct coupons$]$?)

- Let X = time to get n coupons.
- Let X_1 = time to get the first distinct coupon.
  - We have $\mathbb{E}[X_1] = 1$, for obvious reasons.
- Let X_2 = time to get the second distinct coupon, after getting the first.
  - We have $\mathbb{E}[X_2]$ = geometric with distribution $P(\text{get second} \mid \text{got first}) = \frac{n-1}{n}$, hence $\mathbb{E}[X_2] = \frac{n}{n-1}$.

In general,

$$\mathbb{E}[X_j] = \frac{n}{n-(i-1)} = \frac{n}{n+1-i}$$

Summing these all up, we arrive at

$$\mathbb{E}[X] = \mathbb{E}[X_1 + ... + X_n] = \sum_{i}^{n} \mathbb{E}[X_i] \approx n \ln n$$