Discrete Random Variables — Modules

20. Bernoulli, Binomial, Hypergeometric Distrns

21. Geometric, Negative Binomial Distrns

22. Poisson Distrn

3.20 Bernoulli, Binomial, Hypergeometric Distrns

Bernoulli Distrn

Binomial Distrn

Hypergeometric Distrn

Bernoulli(p) Distribution

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } q \end{cases}$$

Previous work showed that $\mathrm{E}[X] = p$, $\mathrm{Var}(X) = pq$, and $M_X(t) = pe^t + q$.

Further, $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ Bern(p) $\Rightarrow \sum_{i=1}^n X_i \sim \text{Bin}(n,p)$.

Binomial(n, p) Distribution

Let $Y = \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ Bern(p). Then $Y \sim \text{Bin}(n, p)$.

$$\Pr(Y = k) \;=\; \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

Example: Toss 2 dice 5 times. Let Y be the number of 7's you see. $Y \sim \text{Bin}(5, 1/6)$. Then, e.g.,

$$\Pr(Y = 4) \;=\; \binom{5}{4} (\frac{1}{6})^4 (\frac{5}{6})^{5-4}.$$

$X_1, \ldots X_n$ i.i.d. Bern(p) implies

$$\mathsf{E}[Y] = \mathsf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathsf{E}[X_i] = np$$

and, similarly,

$$\mathsf{Var}(Y) = npq.$$

We've already seen that $M_Y(t) = (pe^t + q)^n$.

If $Y_1, \ldots, Y_k$ are *indep* and $Y_i \sim \mathsf{Bin}(n_i, p)$, then

$$\sum_{i=1}^{k} Y_i \sim \mathsf{Bin}\Big(\sum_{i=1}^{k} n_i, p\Big).$$

Hypergeometric Distribution

You have a objects of type 1 and b objects of type 2.

Select n objects w/o replacement from the $a+b$.

Let X be the number of type 1's selected.

$$\Pr(X=k) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}, \quad k=0,1,\ldots,n.$$

After some algebra, it turns out that

$$\mathsf{E}[X] \;=\; n\Big(\frac{a}{a+b}\Big) \quad \text{and}$$

$$\mathsf{Var}(X) \;=\; n\Big(\frac{a}{a+b}\Big)\Big(1-\frac{a}{a+b}\Big)\Big(\frac{a+b-n}{a+b-1}\Big).$$

$\frac{a}{a+b}$ here plays the role of p in the Binomial distrn.

Old Example: 25 sox in a box. 15 red, 10 blue. Pick 7 w/o replacement.

$$\Pr(\text{exactly 3 reds are picked}) = \frac{\binom{15}{3}\binom{10}{4}}{\binom{25}{7}}$$

3.21 Geometric, Negative Binomial Distrns

Geometric Distribution

Memoryless Property of Geometric

Negative Binomial Distribution

Comparison of Binomial and Negative Bin

Geometric(p) Distribution

Suppose we consider an infinite sequence of indep Bern(p) trials.

Let Z equal the number of trials *until the first success* is obtained. The event $Z = k$ corresponds to $k - 1$ failures, and then a success. Thus,

$$\Pr(Z = k) \; = \; q^{k-1}p, \quad k = 1, 2, \ldots .$$

Z has the **Geometric(p) distribution**.

The mgf of the Geom(p) is

$$M_Z(t) \;=\; \mathsf{E}[e^{tZ}] \;=\; \sum_{k=1}^{\infty} e^{tk} q^{k-1} p$$

$$=\; pe^t \sum_{k=0}^{\infty} (qe^t)^k$$

$$=\; \frac{pe^t}{1-qe^t}, \text{ for } qe^t < 1.$$

So

$$M_Z(t) \;=\; \frac{pe^t}{1-qe^t}, \text{ for } t < \ell\mathrm{n}(1/q).$$

Thus,

$$
\begin{aligned}
\mathsf{E}[Z] &= \frac{d}{dt} M_Z(t)\Big|_{t=0} \\
&= \frac{(1-qe^t)(pe^t) - (-qe^t)(pe^t)}{(1-qe^t)^2}\Big|_{t=0} \\
&= \frac{pe^t}{(1-qe^t)^2}\Big|_{t=0} \\
&= \frac{p}{(1-q)^2} = \frac{1}{p}.
\end{aligned}
$$

Similarly, after a lot of algebra,

$$\mathsf{E}[Z^2] = \frac{d^2}{dt^2} M_Z(t)\Big|_{t=0} = \frac{2-p}{p^2},$$

so that

$$\mathsf{Var}(Z) = \mathsf{E}[Z^2] - (\mathsf{E}[Z])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Example: Toss a die repeatedly. What's the prob that we observe a '3' for the first time on the 8th toss?

Answer: The number of tosses we need is $Z \sim \text{Geom}(1/6)$. $\Pr(Z = 8) = (5/6)^7(1/6)$.

How many tosses would we expect to take?

Answer: $E[Z] = 1/p = 6$ tosses.

Memoryless Property of Geometric

Theorem: Suppose that $Z \sim \text{Geom}(p)$. Then for positive integers s, t, we have

$$\text{Pr}(Z > s + t | Z > s) \; = \; \text{Pr}(Z > t).$$

Why is it the **memoryless property**? Well, Tommy, if an event hasn't occurred by time s, the prob that it will occur after an additional t time units is the same as the (unconditional) prob that it will occur after time t — it forgot that it made it past time s!

Proof:

$$\Pr(Z > s + t | Z > s)$$

$$= \frac{\Pr(Z > s + t \cap Z > s)}{\Pr(Z > s)}$$

$$= \frac{\Pr(Z > s + t)}{\Pr(Z > s)} \quad (t \text{ positive})$$

$$= \frac{\sum_{j=s+t+1}^{\infty} q^{j-1} p}{\sum_{j=s+1}^{\infty} q^{j-1} p} = \frac{q^{s+t} \sum_{j=0}^{\infty} q^j}{q^s \sum_{j=0}^{\infty} q^j}$$

$$= q^t.$$

Meanwhile,

$$\begin{aligned}
\Pr(Z > t) &= \sum_{j=t+1}^{\infty} q^{j-1} p \\
&= pq^t \sum_{j=0}^{\infty} q^j \\
&= \frac{pq^t}{1-q} \\
&= q^t.
\end{aligned}$$

Thus, $\Pr(Z > s + t | Z > s) = \Pr(Z > t)$. Done.

Fun Fact: The Geom(p) is the only discrete distribution with the memoryless property.

Not-so-Fun Fact: Some books define the Geom(p) as the number of Bern(p) *failures* until you observe a success. # failures $=$ # trials -1. You should be aware of this inconsistency, but don't worry about it for now.

Negative Binomial Distribution (aka Pascal distrn)

Suppose we consider an infinite sequence of indep Bern(p) trials.

Now let Z equal the number of trials *until the* r*th success* is obtained. $Z = r, r+1, \ldots$. The event $Z = k$ corresponds to exactly $r-1$ successes by time $k-1$, and then the rth success at time k.

'FFFFFSFS' corresponds to $Z = 7$ trials until the $r =$ 2nd success.

Notation: $Z \sim$ NegBin(r, p).

Remark: As with the Geom(p), the exact definition of the NegBin depends on what book you're reading.

Theorem: If $Z_1, \ldots, Z_r \stackrel{\text{iid}}{\sim}$ Geom(p), then $Z = \sum_{i=1}^{n} Z_i \sim$ NegBin(r, p).

Proof: Won't do it here, but you can use the mgf technique.

Anyhow, it makes sense if you think of Z_i as the number of trials after the $(i-1)$st success up to and including the ith success.

Since the Z_i's are i.i.d., the above theorem gives:

$$\mathsf{E}[Z] = r\mathsf{E}[Z_i] = r/p,$$

$$\mathsf{Var}(Z) = r\mathsf{Var}(Z_i) = rq/p^2,$$

$$M_Z(t) = [M_{Z_i}(t)]^r = \left(\frac{pe^t}{1-qe^t}\right)^r.$$

Just to be complete, let's get the pmf of Z.

$Z = k$ iff get exactly $r-1$ successes by time $k-1$, and then the rth success at time k. So...

$$\begin{aligned} \Pr(Z = k) &= \left[\binom{k-1}{r-1} p^{r-1} q^{k-r}\right] p, \quad k = r, r+1, \ldots \\ &= \binom{k-1}{r-1} p^r q^{k-r}, \quad k = r, r+1, \ldots. \end{aligned}$$

How are the Bin and NegBin Related?

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p) \;\Rightarrow\; X \equiv \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$.

$Z_1, \ldots, Z_r \overset{\text{iid}}{\sim} \text{Geom}(p) \;\Rightarrow\; Z \equiv \sum_{i=1}^r Z_i \sim \text{NegBin}(r, p)$.

$\text{E}[X] = np$, $\text{Var}(X) = npq$.

$\text{E}[Z] = r/p$, $\text{Var}(Z) = rq/p^2$.

3.22 Poisson Distribution

Poisson Process

Poisson Distribution

Properties

Poisson Process

Let $N(t)$ be a **counting process**. That is, $N(t)$ is the number of occurrences (or arrivals, or events) of some process over the time interval $[0, t]$. $N(t)$ looks like a step function.

Examples: $N(t)$ could be any of the following.
(a) Cars entering a shopping center (time).
(b) Defects on a wire (length).
(c) Raisins in cookie dough (volume).

Let $\lambda > 0$ be the average number of occurrences per unit time (or length or volume).

In the above examples, we might have:

(a) $\lambda = 10$/min. (b) $\lambda = 0.5$/ft. (c) $\lambda = 4/\text{in}^3$.

A Poisson process is a specific counting process. . .

First, some notation: $o(h)$ is a generic function that goes to zero faster than h goes to zero.

Definition: A **Poisson process** is one that satisfies the following assumptions:

(1) There is a short enough interval of time, say of length h, such that, for all t,

$$\begin{aligned} \Pr(N(t+h) - N(t) = 0) &= 1 - \lambda h + o(h) \\ \Pr(N(t+h) - N(t) = 1) &= \lambda h + o(h) \\ \Pr(N(t+h) - N(t) \geq 2) &= o(h) \end{aligned}$$

(2) If $t_1 < t_2 < t_3 < t_4$, then $N(t_4) - N(t_3)$ and $N(t_2) - N(t_1)$ are *indep* RV's.

English translation of Poisson process assumptions.

(1) Arrivals basically occur one-at-a-time, and then at rate λ/unit time. (We must make sure that λ doesn't change over time.)

(2) The numbers of arrivals in two disjoint time intervals are indep.

Poisson Process Example: Neutrinos hit a detector. Occurrences are rare enough so that they really do happen one-at-a-time. You never get arrivals of groups of neutrinos. Further, the rate doesn't vary over time, and all arrivals are indep of each other.

Anti-Example: Customers arrive at a restaurant. They show up in groups, not one-at-a-time. The rate varies over the day (more at dinnertime). Arrivals may not be indep. This ain't a Poisson process.

Poisson Distribution

Definition: Let X be the number of occurrences in a Poisson(λ) process in a *unit interval* of time. Then X has the **Poisson distribution** with parameter λ.

Notation: $X \sim$ Pois(λ).

Theorem/Definition: $X \sim$ Pois(λ) $\Rightarrow$
$\Pr(X = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$.

Remark: The value of λ can be changed simply by changing the units of time.

Example:

$X = \#$ calls to a switchboard in 1 minute $\sim$ Pois(3)

$Y = \#$ calls to a switchboard in 5 minutes $\sim$ Pois(15)

$Z = \#$ calls to a switchboard in 10 sec $\sim$ Pois(0.5)

Properties

Theorem: $X \sim \text{Pois}(\lambda) \Rightarrow$ mgf is $M_X(t) = e^{\lambda(e^t - 1)}$.

Proof:

$$
\begin{aligned}
M_X(t) &= \mathsf{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} \left(\frac{e^{-\lambda} \lambda^k}{k!} \right) \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\
&= e^{-\lambda} e^{\lambda e^t}.
\end{aligned}
$$

Theorem: $X \sim \text{Pois}(\lambda) \Rightarrow \text{E}[X] = \text{Var}(X) = \lambda$.

Proof (using mgf):

$$\begin{aligned}
\text{E}[X] &= \frac{d}{dt} M_X(t)\Big|_{t=0} \\
&= \frac{d}{dt} e^{\lambda(e^t - 1)}\Big|_{t=0} \\
&= \lambda e^t M_X(t)\Big|_{t=0} \quad \text{(chain rule)} \\
&= \lambda \quad \text{(after algebra)}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathsf{E}[X^2] &= \frac{d^2}{dt^2}M_X(t)\Big|_{t=0} = \frac{d}{dt}\Big(\frac{d}{dt}M_X(t)\Big)\Big|_{t=0} \\
&= \lambda\frac{d}{dt}(e^t M_X(t))\Big|_{t=0} \\
&= \lambda\Big[e^t M_X(t) + e^t\frac{d}{dt}M_X(t)\Big]\Big|_{t=0} \\
&= \lambda e^t\Big[M_X(t) + \lambda e^t M_X(t)\Big]\Big|_{t=0} \\
&= \lambda(1+\lambda).
\end{aligned}$$

Thus,

$$\text{Var}(X) = \text{E}[X^2] - (\text{E}[X])^2 = \lambda(1+\lambda) - \lambda^2 = \lambda.$$

Done.

Example: Calls to a switchboard arrive as a Poisson process with rate 3 calls/min.

Let $X =$ number of calls in 40 sec. So $X \sim \text{Pois}(2)$.

$\text{E}[X] = \text{Var}(X) = 2$, $\Pr(X \leq 3) = \sum_{k=0}^{3} e^{-2}2^k/k!$

Theorem (Additive Property of Poissons): Suppose $X_1, \ldots, X_n$ are *indep* with $X_i \sim$ Pois(λ_i), $i = 1, \ldots, n$. Then

$$Y \equiv \sum_{i=1}^{n} X_i \sim \text{Pois}\Big(\sum_{i=1}^{n} \lambda_i\Big).$$

Proof:

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \quad (X_i\text{'s indep}) \\ &= \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{(\sum_{i=1}^{n} \lambda_i)(e^t-1)}, \end{aligned}$$

which is the mgf of the Pois$(\sum_{i=1}^{n} \lambda_i)$ distribution.