May 9 Lec—Poisson Cont., Moment Generating Functions, Mgfs of Binomial & Geometric

Proposition

If $X \sim P(\lambda)$ then

1. $\mathbb{E})X = \lambda$
2. $V(X) = \lambda$

Proof

1. $$\mathbb{E}(X) = \sum_{k=0}^{\infty} k p_X(k)$$
$$= \sum_{k=0}^{\infty} k e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$
$$= e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!}$$
$$= e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$
Change variable: $l = k - 1$
$$= e^{-\lambda} \cdot \sum_{l=0}^{\infty} \frac{\lambda^{l+1}}{l!}$$
$$= e^{-\lambda} \cdot \lambda \cdot \sum_{k=0}^{\infty} \frac{\lambda}{l!}$$
$$= e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda$$

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

Computing $\mathbb{E}(X(X-1))$:

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k \cdot (k-1) \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$= \sum_{k=2}^{\infty} k \cdot (k-1) \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}$$

Change variables, $l = k-2$:

$$= e^{-\lambda} \sum_{l=0}^{\infty} \frac{\lambda^{k+2}}{l!}$$

$$= e^{-l} \cdot \lambda^2 \cdot e^{l} = \lambda^2$$

So

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Exercise

$\mathbb{E}(X(X-1)(X-2)) = \lambda^3$
$\mathbb{E}(X(X-1)(X-2)\ldots(X-r)) = \lambda^r$

For a geometric random variable $X \sim G(p)$, the expected value is $\frac{1}{p}$ and the most probable (modal) value is 1. But it's not like this with the Poisson.

Example

See 3.143
Let X be the number of calls to a fire department in a given day, so $X \sim P(5.3)$. What is the most likely number of calls received by the fire department?
We are looking to maximize $a_k := \frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$. You might have thought about differentiating this expression, but this doesn't make sense because k is discrete.

Let's take the ratio $\frac{a_{k+1}}{a_k}$, which is > 1 if the sequence is increasing and < 1 if the sequence is decreasing.
We calculate

$$\frac{a_{k+1}}{a_k} = \frac{\lambda^{k+1}}{(k+1)!} \cdot \frac{k!}{\lambda^k} = \frac{\lambda}{k+1}$$

We solve for

$$\frac{\lambda}{k+1} < 1$$
$$\implies k+1 > \lambda = 5.3$$
$$\implies k > 4.3$$

So for $k \geq 5$ we have $a_{k+1} < a_k$, or that the sequence is decreasing from 5 to ∞.
Similarly we have

$$\frac{a_{k+1}}{a_k} = \frac{\lambda}{k+1} > 1 \iff k < 4.3$$

so the sequence is increasing from 0 to 5.
This shows that the k with the highest $p_X(k)$ is $k = 5$.

Turns out (professor does not provide proof) that the most likely (modal) value is $\lfloor \lambda \rfloor$, i.e. round λ down to the nearest natural number. If $\lambda \in \mathbb{N}$ then (professor thinks) that λ and $\lambda + 1$ has equal probability.

Example

Continuing from example above.
What is the probability of receiving at least 1 call?

$$P(X \geq 1) = 1 - P(X < 1)$$
$$= 1 - P(X = 0)$$
$$= 1 - e^{-}\lambda = 1 - e^{-5.3}$$

2nd part:
Assume that the operating cost for that fire station on a given day is $Y = 2^X$. Find the expected cost.

Well, if you don't know what to do in math, you go back to the definition!
$\mathbb{E}(Y) = \sum_{k=0}^{\infty} 2^k \cdot p_X(k)$

$$= \sum_{k=0}^{\infty} 2^k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!}$$

$$= e^{-\lambda} \cdot e^{2\lambda} = e^{\lambda} = e^{5.3}$$

Exercise

Suppose that there is a garage with two entries 1 and 2. Let X_1 be the number of cars that come in through Entry 1, and X_2 be the number of cars that come in through Entry 2. Assume (reasonably) $X_1 \sim P(\lambda_1)$ and $X_2 \sim P(\lambda_2)$. Also assume that X_1 and X_2 are independent, i.e. the events $\{X_1 = k_1\}$ and $\{X_2 = k_2\}$ are independent $\forall k_1, k_2$.
Consider the random variable $X = X_1 + X_2$. Prove that $X \sim P(\lambda_1 + \lambda_2)$.

The professor offers the first several steps to the proof:
$X \in \{0, 1, 2, ...\}$ is clear.
For the probability function,

$$p_X(0) = P(X = 0)$$

$$= P(X_1 + X_2 = 0)$$

$$= P(\{X_1 = 0\} \cap X_2 = 0)$$

$$= P(\{X_1 = 0\}) \cdot P(\{X_2 = 0\})$$

$$= e^{-\lambda_1} \cdot e^{-\lambda_2} = e^{-(\lambda_1 + \lambda_2)}$$

$$p_X(1) = P(X = 1)$$

$$= P(X_1 + X_2 = 1)$$

$$= P(\{X_1 = 1\} \cap \{X_2 = 0\}) + P(\{X_1 = 0\} \cap \{X_2 = 1\})$$

$$= P(X_1 = 1) \cdot P(X_2 = 0) + P(X_1 = 0) \cdot P(X_2 = 1)$$

$$= \lambda_1 e^{-\lambda_1} e^{-\lambda_2} + e^{-\lambda_1} \lambda_2 e^{-\lambda_2}$$
$$= (\lambda_1 + \lambda_2) e^{-(\lambda_1+\lambda_2)}$$

Fill in the gap between

$$P(X=n) = \sum_{k=0}^{\infty} P(X_1 = k) P(X_2 = n-k)$$

and

$$P(X=n) = \frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)}$$

Moment generating function

i.e. more Laplace transforms 😭😭😭

Definition: Moment generating function

Given a (discrete) random varibale X, the moment generating function (abbr. mgf) of X is the function defined as

$$m_X(t) = \mathbb{E}(e^{tx})$$

Remarks

1. The domain of the function m_X may or may not be equal to R, but (remark 2)
2. $m_x(0) = \mathbb{E}(e^{0x}) = \mathbb{E}(1) = 1$, so $0 \in Dom(m_X)$ (domain of m_X always contains the point 0)
3. $Dom(m_X)$ is always connected, so it is always an interval which contains 0.

Definition

Given a (discrete) random variable X, the nth moment of X is defined as $\mu_n = \mathbb{E}(X^n)$.

Example

Suppose we have the following information:

$p_X(-2) = \frac{1}{2}$
$p_X(0) = \frac{1}{3}$
$p_X(1) = \frac{1}{6}$

We have

$$\begin{aligned} m_X(t) &= \mathbb{E}(e^{tx}) \\ &= \sum_x e^{tx} p_X(x) \\ &= e^{-2t} \cdot \frac{1}{2} + e^{0t} \cdot \frac{1}{3} + e^{t} \cdot \frac{1}{6} \\ &= \frac{1}{2}e^{-2t} + \frac{1}{6}e^{t} + \frac{1}{3} \end{aligned}$$

(For those of you people shouting that's a sum of transformations of Dirac functions, you're right)

Borrowing a theorem from ODEs (*feeling Prof Humphries's eyes on me*):

Theorem

The moment generating function is a unique identifier of the probability function of a random variable. Put more mathematically: Denote by P_I the set of all probability functions over a connected subset $I \subseteq \mathbb{R}$, and P_I' the set of all Laplace transforms of P_I, the function $\phi : P_I \to P_I'$, $f \mapsto \mathcal{L}\{f(x)\}$ is injective.

Example

$$m_X(t) = \frac{1}{4}e^{-3t} + \frac{1}{3}e^{4t} + \frac{1}{4} + \frac{1}{6}e^{-t}$$

We have $P(X=6)=0$, and $P(X=0)=\frac{1}{4}$.

"Linearity" of mgfs

If $Y = aX + b$, then $m_Y(t) = e^{bt}m_X(at)$.

Proof

$$m_Y(t) = \mathbb{E}(e^{ty})$$

$$= \mathbb{E}(e^{(ax+b)t})$$

$$= \mathbb{E}(e^{atx} \cdot e^{bt})$$

$$= e^{bt}\mathbb{E}(e^{atx})$$

$$= e^{bt} \cdot m_X(at)$$

What is the connection between m_X and $\mathbb{E}(X^n) = \mu_n$?
We have

$$m_X(t) = \mathbb{E}(e^{tx})$$

$$= \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(tx)^n}{n!}\right)$$

$$= \mathbb{E}\left(\sum_{n=0}^{\infty} x^n \frac{t^n}{n!}\right)$$

By (???) Convergence Theorem,

$$= \sum_{n=0}^{\infty} \mathbb{E}(x^n) \cdot \frac{t^n}{n!}$$

Recall

For a function $g(t)=\sum_{n=0}^{\infty}a_nt^n$ $\forall t$ such that $|t|<R$, we have $g^{(n)}(0)=n!\cdot a_n$.

So we have

$$\frac{d^n}{dt^n}m_X(t)|_{t=0}=\mathbb{E}(X^n)$$

Now we see see why the mgf is called this.

Theorem

If the domain of m_X contains an internval centered at $t=0$, then $\frac{d^n}{dt^n}|_{t=0}=\mathbb{E}(X^n)$

Example

$p_X(-2)=\frac{1}{2}$
$p_X(0)=\frac{1}{3}$
$p_X(1)=\frac{1}{6}$

We already saw that

$$m_X(t)=\frac{1}{2}e^{-2t}+\frac{1}{3}+\frac{1}{6}e^t \quad \forall t\in\mathbb{R}$$

We compute $\mathbb{E}(X)$ two different ways, using the theorem and using the definition:
$\frac{d}{dt}m_X(t)|_{t=0}=-e^{-2t}+\frac{1}{6}e^t|_{t=0}=-\frac{5}{6}$
Indeed, $\mathbb{E}(X)=-2\cdot\frac{1}{2}+0\cdot\frac{1}{3}+1\cdot\frac{1}{6}=-\frac{5}{6}$
We do the same for $\mathbb{E}(X^2)$:
$\frac{d^2}{dt^2}(m_X(t))=2e^{-2t}\frac{1}{6}e^t|_{t=0}=\frac{13}{6}$
similarly $\mu_2=\mathbb{E}(X^2)=(-2)^2\cdot\frac{1}{2}+0^2\cdot\frac{1}{3}+1^2\cdot\frac{1}{6}=\frac{13}{6}$.

Now let's compute the mgfs of some of the distributions we have already seen.
Binomial:
For $X\sim B(N,p)$ we have $p_X(k)=C_k^Np^k(1-p)^{N-k}$ for $k=0,1,2,\ldots,N$.

Then,

$$m_X(t) = \mathbb{E}(e^{tx})$$

$$= \sum_{k=0}^{N} e^{tk} C_k^N p^k (1-p)^{N-k}$$

$$= \sum_{k=0}^{N} C_k^N (pe^t)^k (1-p)^{N-k}$$

$$= [pe^t + 1 - p]^N$$

The domain is obviously $\forall t \in \mathbb{R}$.

Application

Given that

$$m_X(t) = \left(\frac{e^{2t}}{3} + \frac{2}{3}\right)^{10} e^{-t}$$

for all $t \in \mathbb{R}$, find $V(X)$ and $P(X = 11)$.
By the linearity of mgf, $X = 2Y - 1$ for $m_Y(s) = \left(\frac{e^s}{3} + \frac{2}{3}\right)^{10}$
This implies $Y \sim B\left(10, \frac{1}{3}\right)$.
We have $V(X) = 4V(Y) = 4 \cdot \left(\frac{1}{3} \cdot \frac{2}{3} \cdot 10\right) = \frac{80}{9}$
Also, $P(X = 11) = P(2Y - 1 = 11) = P(2Y = 12) = P(Y = 6) = C_6^{10} \cdot \left(\frac{1}{3}\right)^6 \cdot \left(\frac{2}{3}\right)^4$

Geometric distribution:
$X \sim G(p)$
$p_X(k) = P(X = k) = p(1-p)^{k-1} \ \forall k \in \mathbb{N}^*$.
We do a bunch of calculations:

$$m_X(t) = \mathbb{E}(e^{tx})$$

$$= \sum_{k=1}^{\infty} e^{tk} p (1-p)^{k-1}$$

$$= p \cdot e^t \cdot \sum_{k=1}^{\infty} (e^t \cdot (1-p))^{k-1}$$

$$= pe^t \cdot \frac{1}{1 - e^t(1-p)}$$

$$= \frac{pe^t}{1-(1-p)e^t}$$

Domain: $|(1-p)e^t| < 1 \implies t < -\ln(1-p)$

Application

Given that $m_Y(t) = \frac{e^{4t}}{4-3e^{2t}}$ for $t < -\frac{1}{2}\ln\left(\frac{3}{4}\right)$, find $V(Y)$ and $P(Y = 10)$.
We can brute force $m_Y(t)$ a bit:

$$m_Y(t) = e^{2t} \cdot \frac{\frac{e^{2t}}{4}}{1 - \frac{3}{4}e^{2t}}$$

Again, by the linearity of mgf, $m_Y(t) = e^{2t} m_X(2t) \implies Y = 2X + 2$ where

$$m_X(s) = \frac{e^s}{4 - 3e^s} = \frac{\frac{1}{4}e^s}{1 - \frac{3}{4}e^s}$$

So $X \sim G\left(\frac{1}{4}\right)$.
$V(Y) = 4V(X) = 4 \cdot 4(4-1) = 48$
$P(Y = 10) = P(X = 4) = \left(\frac{3}{4}\right)^3 \cdot \frac{1}{4} = \frac{27}{256}$

Another example

What is $\mathbb{E}(a^x)$ given $m_X(t)$?
$\mathbb{E}(a^x) = \mathbb{E}(e^{\ln(a)\cdot x}) = m_X(\ln(a))$.