Stat 2911 Lecture Notes

Class 6, 2017

Uri Keich

Marginal conditional distribution (binomial, Poisson), Expectation (and L^1) definition and examples (Bernoulli, Poisson, binomial, geometric), Variance definition and calculation

3) $X \sim \text{binom}(m,p)$ is ind. of $Y \sim \text{binom}(n-m,p)$.

We saw $X+Y \sim \text{binom}(n,p)$.

Consider the <u>conditional</u> dist. of X given $X+Y=r$:

$$P(X=k \mid X+Y=r) = \frac{P(X=k, X+Y=r)}{P(X+Y=r)}$$

$$= \frac{P(X=k, Y=r-k)}{P(X+Y=r)}$$

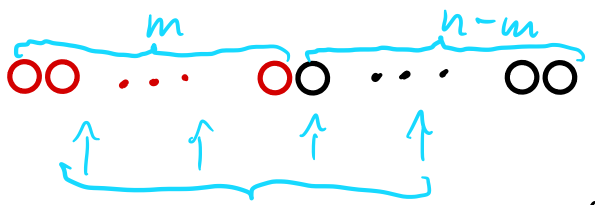
$$= \frac{\binom{m}{k} p^k (1-p)^{m-k} \binom{n-m}{r-k} p^{r-k} (1-p)^{(n-m)-(r-k)}}{\binom{n}{r} p^r (1-p)^{n-r}}$$

$$= \frac{\binom{m}{k}\binom{n-m}{r-k}}{\binom{n}{r}}$$

$\Rightarrow$ Conditioned on $X+Y=r$, $X \sim \text{hyper}(\underset{\text{red}}{m}, \underset{\text{black}}{n-m}, \underset{\text{sample}}{r})$

Why?

Choose r balls (successes) from n and check how many are red (from the first m, "X" trials).



trials

randomly choose r trials (successes)

4) $X \sim \text{Pois}(\lambda)$ ind. of $Y \sim \text{Pois}(\gamma)$

We saw $X+Y \sim \text{Pois}(\lambda+\gamma)$.

Consider the <u>conditional</u> dist. of X given $X+Y=n$.

For $k \in \{0,1,\dots,n\}$,

$$P(X=k \mid X+Y=n) = \frac{P(X=k, Y=n-k)}{P(X+Y=n)}$$

$$= \frac{e^{-\lambda}\lambda^k/k! \; e^{-\gamma}\gamma^{n-k}/(n-k)!}{e^{-(\lambda+\gamma)}(\lambda+\gamma)^n/n!}$$

$$= \binom{n}{k}\left(\frac{\lambda}{\lambda+\gamma}\right)^k\left(\frac{\gamma}{\lambda+\gamma}\right)^{n-k}$$

$$\Rightarrow X \mid X+Y=n \sim \text{binom}\left(n, \frac{\lambda}{\lambda+\gamma}\right)$$

Why?

Imagine two sources emitting particles at rates λ and γ.

Each of the observed particles can be independently attributed to the first source with a fixed prob. related to its relative intensity.

Expectation

To answer all statistical questions about a RV X we need to know p_X, but sometime a summary would suffice.

Def. The expected value, or expectation of, a discrete non-negative RV X with range $X(\Omega) = \{x_k\}$ is

$$E(X) = \sum_k x_k \, p_X(x_k)$$

Examples.

1) $X \sim \text{Bernoulli}(p)$

$$E(X) = 0 \cdot (1-p) + p = \quad \text{(unattainable value)}$$

2) $X \sim \text{Pois}(\lambda)$

$$E(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \lambda^k / k!$$

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

$m = k-1$

$$= \lambda e^{-\lambda} \sum_{m=0}^{\infty} \lambda^m / m!$$

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

3) $X \sim \text{binom}(n,p)$

$$E(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$m = k-1$

$$= np \sum_{m=0}^{n-1} \binom{n-1}{m} p^m (1-p)^{(n-1)-m}$$

$$= np[p+(1-p)]^{n-1} = np$$

4) $X \sim \text{Geometric}(p)$ $\quad p > 0$

$$E(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1} \cdot p$$

$P(X=+\infty)=0$ and the convention here is $\infty \cdot 0 = 0$.

Can this be simplified?

Need $\sum_{k=1}^{\infty} kq^{k-1}$ where $q = 1-p$

Consider the power series

$$f(q) = \sum_{k=0}^{\infty} q^k = \qquad \forall q \in (-1,1)$$

What's the relation between $f(q)$ and $\sum kq^{k-1}$?

Caution. If f_n are differentiable functions then $\frac{d}{dx}\sum_1^N f_n(x) = \sum_1^N f_n'(x)$ but generally we can't assume $\frac{d}{dx}\sum_1^\infty f_n(x) = \sum_1^\infty f_n'(x)$.

However, a power series can be differentiated term by term within its radius of convergence. Therefore,

$$f'(q) = \sum_{k=0}^{\infty} k q^{k-1} \qquad \forall q \in (-1, 1)$$

But

$$f'(q) = \frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{1}{(1-q)^2}$$

$$\Rightarrow \quad \sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$$

$$\Rightarrow \quad E(X) = p\sum_{k=1}^{\infty} k q^{k-1} = p\frac{1}{(1-q)^2} = \frac{1}{p}$$

This makes sense intuitively: when rolling a fair die we expect to roll it 6 times before seeing a "5".

For a negative RV $X \leq 0$ we also define

$$E(X) = \sum_k x_k p_X(x_k),$$

but what about RVs which are both negative and positive?

Using the above definition can be problematic:

let $a_n = (-1)^{n+1}\frac{1}{n}$, then $\sum_n a_n = \log(2)$ but $\sum |a_n| =$

i.e., the series converges conditionally.

Since $a_n \xrightarrow[n]{} 0$ it follows that by rearranging the order of the terms, the series can converge to <u>any</u> finite number!

By analogy, $\sum_k x_k p_X(x_k)$ might converge to different numbers depending on the order of $\{x_k\}$ which is arbitrary.

Therefore we need to make sure the expectation does not depend on the order we sum our elements.

We denote by $L^1 = L^1(\Omega, \mathcal{F}, P) = L^1(\Omega)$, the space of all discrete RVs X s.t.

$$\left\{ X : \Omega \to \mathbb{R} \text{ is a discrete RV with } \sum_{x \in X(\Omega)} |x| p_X(x) < \infty \right\}$$

Exercise, (i) If $|X(\omega)| \leq C < \infty \ \forall \omega \in \Omega$ then $X \in L^1$

(ii) If X is Bernoulli, binomial, geometric, Pois $\Rightarrow X \in L^1$

(iii) If $X \geq 0$ then $X \in L^1 \iff E(X) < \infty$

Def. For $X \in L^1$ with $X(\Omega) = \{x_k\}$ we define its expectation by
$$E(X) = \sum_k{}' \, x_k \, p_X(x_k)$$

Claim. If $X \in L^1$ then the series $\sum' x_k p_X(x_k) \in \mathbb{R}$ does not depend on the order of summation.

Proof. Let $a_k = x_k p_X(x_k)$ then $\sum' |a_k| = \sum' |x_k| p_X(x_k) < \infty$. That is, the series $\sum' a_k$ converges absolutely (or is absolutely convergent) and such a series does not depend on the order of summation. Finally,
$$\left| \sum' x_k p_X(x_k) \right| \overset{2}{\leq} \sum' |x_k| \, p_X(x_k) < \infty .$$

Cor. $E(X)$ is well-defined for $X \in L^1$ (as well as for $X \geq 0$ and $X \leq 0$).

Expectation of a function of a RV

Let X be a discrete RV with range $X(\Omega) = \{x_k\}$, and let $g : \mathbb{R} \to \mathbb{R}$.
Then $Y = g(X)$ is a discrete RV with pmf p_Y.
Deciding whether $Y \in L^1$ and, if so, computing $E(Y)$ should be done using p_Y. However, the following claim shows we can determine those things without finding p_Y.

<u>Claim.</u> (i) $Y \in L^1$ iff $\sum_k{}' |g(x_k)| \, p_{\mathbb{X}}(x_k) < \infty$

(ii) If $Y \in L^1$ then $E(Y) = \sum_k{}' g(x_k) \, p_{\mathbb{X}}(x_k)$

<u>Proof.</u> (later)

<u>Cor.</u> If $g \geq 0$ then $E(Y) = \sum_k{}' g(x_k) \, p_{\mathbb{X}}(x_k)$

Example. $g(x) = x^2$, $E(\mathbb{X}^2) = \sum_k{}' x_k^2 \, p_{\mathbb{X}}(x_k)$

<u>Proof</u> (of Cor) (skipped in class)

$Y \geq 0$ so, $E(Y)$ is well defined regardless of whether $Y \in L^1$.

By (ii) of the claim $Y \in L^1 \Rightarrow E(Y) = \sum_i{}' g(x_k) \, p_{\mathbb{X}}(x_k)$,

By (i), if $Y \notin L^1$ then $\sum_i{}' |g(x_k)| \, p_{\mathbb{X}}(x_k) = +\infty$

$$\Rightarrow \sum{}' g(x_k) \, p_{\mathbb{X}}(x_k) = +\infty$$

but $Y \notin L^1 \Rightarrow \sum_i{}' |y_n| \, p_Y(y_n) = +\infty$

$$\Rightarrow E(Y) = \sum_i{}' y_n \, p_Y(y_n) = +\infty$$

So, $E(Y) = \sum_i{}' g(x_k) \, p_{\mathbb{X}}(x_k)$ in this case as well. $\square$

Variance (measures the dispersion about the mean)

Suppose $\underline{X} \in L^1$ and denote $\mu = E(\underline{X})$.

How would you measure the dispersion about μ?

Def. The *variance* of $\underline{X}$ is defined as

$$Var(\underline{X}) = E[\underbrace{(\underline{X}-\mu)^2}_{\geq 0}] \qquad (\text{also } V(\underline{X}))$$

Computing $V(\underline{X})$

Let $g(x) = (x-\mu)^2$, then $g(x) \geq 0 \ \forall x$, so (why?)

$$Var(\underline{X}) = \sum_k{}' (x_k-\mu)^2 p_{\underline{X}}(x)$$

Moreover,

$$\sum{}' (x_k-\mu)^2 p_{\underline{X}}(x_k) = \sum{}' (x_k^2 - 2\mu x_k + \mu^2) p(x_k)$$

$$\overset{?}{=} \sum{}' x_k^2 p(x_k) - 2\mu \sum{}' x_k p(x_k) + \mu^2 \sum{}' p(x_k)$$

$$\overset{?}{=} E(\underline{X}^2) - 2\mu E(\underline{X}) + \mu^2$$

$$= E(\underline{X}^2) - \underbrace{[E(\underline{X})]^2}_{\text{also } E^2(\underline{X})}$$

Cor $V(\underline{X}) = E(\underline{X}^2) - E^2(\underline{X})$.

Warning: $\sum_1^N{}' (a_n + b_n) = \sum_1^N{}' a_n + \sum_1^N{}' b_n$

But take $a_n = \frac{1}{n} + \frac{1}{n^2}$ $\quad b_n = -\frac{1}{n}$, then $a_n + b_n = \frac{1}{n^2}$

$$\pi^2/6 = \sum_1^\infty{}' (a_n + b_n) \neq \sum_1^\infty{}' a_n + \sum_1^\infty{}' b_n = \text{"}\infty - \infty\text{"}$$