Stat 2911 Lecture Notes

Class 5, 2017

Uri Keich

Poisson approximation to the binomial distribution, faulty monitor, sum of independent RVs (convolution): binomial, Poisson RVs

The Poisson Dist.

$p_X(k) = e^{-\lambda} \lambda^k / k!$ $\quad k \in \mathbb{N} \quad \lambda > 0$

Models the # of events in a given time interval, assuming:

1) The dist. of the number of events at any time interval depends only on the interval's length/duration.
2) The # of events recorded in two disjoint time intervals are ind. of one another.
3) No two events occur at the same time point.

Let $X = X_1 =$ # of events in $(0,1]$.

We showed $P(X=0) = e^{-\lambda} = e^{-\lambda} \lambda^0/0!$ by showing

$P(X_t = 0) = e^{-\lambda t}$ for some $\lambda > 0$.

To study $P(X=j)$ for $j \geq 1$ we look at

$Y_n =$ # of intervals $(\frac{k-1}{n}, \frac{k}{n}]$ in $(0,1]$ in which an event occurred.

$Y_n \sim \text{Binomial}(n, p_n = 1 - e^{-\lambda/n})$ with $\lambda > 0$

$Y_n \leq X$, but $\lim_{n \to \infty} Y_n(\omega) = X(\omega)$

intuitively

$$\Longrightarrow \quad P(Y_n = k) \xrightarrow[n]{} P(X = k)$$

Note that $Y_n(\omega) \longrightarrow X(\omega)$ even if $X(\omega) = +\infty$ and that this analysis shows $P(X = +\infty) = 0$.

The expected value of a Binomial(n,p) RV is np.

For Y_n this is $np_n = n(1-e^{-\lambda/n})$. (expected # of "$1/n$ intervals" with ≥ 1 events)

What is $\lim np_n$?

First order Taylor series expansion of e^x:

$$e^x = 1 + x + R_1(x) \quad \text{where} \quad R_1(x)/x \xrightarrow[x\to 0]{} 0$$

$$\begin{aligned} np_n &= n(1-e^{-\lambda/n}) \\ &= n\left[1-\left(1-\lambda/n + R_1(-\lambda/n)\right)\right] \\ &= \lambda + \frac{R_1(-\lambda/n)}{-\lambda/n}\cdot\lambda \xrightarrow[n\to\infty]{} \end{aligned}$$

$(x = -\lambda/n)$

<u>Claim</u>. If $Y_n \sim \text{Binom}(n, p_n)$ s.t. $np_n \xrightarrow[n]{} \lambda > 0$, then for any fixed $k \in \mathbb{N}$,

$$P(Y_n = k) \xrightarrow[n]{} e^{-\lambda}\lambda^k/k!.$$

<u>Cor</u>. $X = X_1 \sim \text{Poisson}(\lambda)$: $p_X(k) = e^{-\lambda}\lambda^k/k!$ $\quad k \in \mathbb{N}$.

Proof of Cor.

$\forall k \in \mathbb{N}$, $\quad P(X_1 = k) = \lim_n P(Y_n = k) = e^{-\lambda}\lambda^k/k!$.

<u>Proof of Claim.</u> $P(Y_n=k) = \overbrace{\binom{n}{k} p_n^k}^{\mathrm{I}} \overbrace{(1-p_n)^{n-k}}^{\mathrm{II}}$ $\quad k=0,1,\dots,n$

(I) $$\binom{n}{k} p_n^k = \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^k$$
$$= \frac{(np_n)^k}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \xrightarrow[n\to\infty]{}$$

(II) $$(1-p_n)^{n-k} = (1-p_n)^n (1-p_n)^{-k}$$

Because $np_n \xrightarrow[n]{} \lambda \in \mathbb{R}$ it follows that $p_n \rightarrow$.

Hence $(1-p_n)^{-k} \xrightarrow[n]{}$

$$(1-p_n)^n = e^{n \log(1-p_n)}$$

Taylor expansion: $\log(1-x) = -x + R_1(x)$ where $\frac{R_1(x)}{x} \xrightarrow[x\to 0]{} 0$

$$\Rightarrow n\log(1-p_n) = n[-p_n + R_1(p_n)]$$
$$= -np_n + \frac{R_1(p_n)}{p_n} \cdot (np_n) \xrightarrow[n]{}$$

$$\Rightarrow \lim_n (1-p_n)^n = \lim_n e^{n\log(1-p_n)}$$
$$\overset{?}{=} e^{\lim n \log(1-p_n)}$$
$$= e^{-\lambda}$$

$$\Rightarrow \lim_n P(Y_n = k) = e^{-\lambda} \lambda^k / k! \qquad \square$$

Example. Poisson approx. to the binomial dist.!

$n = 100$, $p = 0.01$ $\lambda = np = 1$

$Y_n \sim Binom(n,p)$ $P(Y_n = 3) \approx 0.0610$

$X_\lambda \sim Poisson(\lambda)$ $P(X_\lambda = 3) \approx 0.0613$

Going back to Bortkiewicz:

# of horse kick fatalities in 1 corp year

20 corps × 10 years = 200 Corp years

# of deaths	count	observed frequency	Poisson approx.
0	109	0.545	0.543
1	65	0.325	0.331
2	22	0.110	0.101
3	3	0.015	0.021
4	1	0.005	0.003

What is missing?

How was λ determined?

$$\text{Empirical mean} = \frac{109}{200} \times 0 + \frac{65}{200} \times 1 + \frac{22}{200} \times 2 + \frac{3}{200} \times 3 + \frac{1}{200} \times 4$$

$$= 0.61$$

For a Poisson RV X, $E(X) = \lambda \Rightarrow \hat{\lambda} = 0.61$.

More on that later!

Example. Faulty monitor (Rice p. 88)

radioactive source

α particles $\longrightarrow$ emitted according to a Poisson(λ) dist.

faulty monitor

registers each incoming particle with prob. p independently of all other particles.

Let X = # of <u>registered</u> particles in an hour.

What is the dist. of X?

Let N = # of <u>emitted</u> particles in the same hour (unobserved).

$N \sim$

For $k \in \mathbb{N}$,

$$P(X=k) \overset{\ell}{=} \sum_{n=0}^{\infty} P(X=k \mid N=n)\, P(N=n)$$

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

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

$$\overset{m=n-k}{=} \frac{e^{-\lambda}(\lambda p)^k}{k!} \sum_{m=0}^{\infty} \frac{[\lambda(1-p)]^m}{m!}$$

$$= \frac{e^{-\lambda}(\lambda p)^k}{k!} e^{\lambda(1-p)} = e^{-\lambda p}\frac{(\lambda p)^k}{k!}$$

$\Longrightarrow X \sim \text{Poisson}(\lambda p)$

It makes sense: our three assumptions still hold!

Moreover, recalling the expected # of counts is the Poisson parameter, it makes sense that $\lambda \longrightarrow \lambda p$.

Dist. of a sum of RVs (arithmetics of RVs)

If X and Y are jointly-distributed (defined on the same sample space) $\mathbb{N}$-valued RVs then $Z = X + Y$ is also an $\mathbb{N}$-valued RV, and for $n \in \mathbb{N}$,

$$
\begin{aligned}
p_Z(n) &= P(Z = n) \\
&= P(X + Y = n) \\
&= \sum_{k=0}^{n} P(X = k, Y = n-k) \\
&= \sum_{k=0}^{n} p_{XY}(k, n-k)
\end{aligned}
$$

If X and Y are ind., then

$$p_Z(n) = \sum_{k=0}^{n} p_X(k)\, p_Y(n-k) \qquad \forall n \in \mathbb{N}$$

p_Z is the convolution of p_X and p_Y denoted as

$$p_Z \equiv p_X * p_Y .$$

Examples

1) $X \sim$ binomial (m,p) is ind. of $Y \sim$ binom$(n-m,p)$.

Let $Z = X + Y$.

It is clear that $Z \sim$.

Formally, for $\ell \in \{0,1,\dots,n\}$

$$P(Z=\ell) = \sum_{k=0}^{\ell} \binom{m}{k} p^k (1-p)^{m-k} \binom{n-m}{\ell-k} p^{\ell-k} (1-p)^{n-m-(\ell-k)}$$

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

Claim. $\sum_{k=0}^{\ell} \binom{m}{k}\binom{n-m}{\ell-k} = \binom{n}{\ell}$

Cor. $P(Z=\ell) = \binom{n}{\ell} p^\ell (1-p)^{n-\ell}$ or, $Z \sim$ binom(n,p)

Proof (of Claim).

(I) Combinatorics : count the same object twice

$\binom{n}{\ell} =$ # of committees of ℓ students from a class of n

Suppose the class has m boys and $n-m$ girls, then the same number can be counted by

$$\sum_{k=0}^{\ell} \text{# of the committees with } k \text{ boys} = \sum_{k=0}^{\ell} \binom{m}{k}\binom{n-m}{\ell-k}$$

(II) Algebra: $(1+x)^n = (1+x)^m (1+x)^{n-m}$

Consider the coefficient of x^l on both sides

LHS: $\binom{n}{l}$ (binomial theorem)

RHS: $\sum_{k=0}^{l} [\text{coeff of } x^k \text{ in } (1+x)^m] \cdot [\text{coeff of } x^{l-k} \text{ in } (1+x)^{n-m}]$

$$= \sum_{k=0}^{l} \binom{m}{k}\binom{n-m}{l-k}$$

2) $X \sim \text{Poisson}(\lambda)$ ind. of $Y \sim \text{Poisson}(\gamma)$.

What is the dist. of $Z = X + Y$? For $n \in \mathbb{N}$,

$$P(Z=n) = \sum_{k=0}^{n} p_X(k)\, p_Y(n-k)$$

$$= \sum_{k=0}^{n} e^{-\lambda} \lambda^k / k! \; e^{-\gamma} \gamma^{n-k} / (n-k)!$$

$$= e^{-(\lambda+\gamma)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \gamma^{n-k}$$

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

$\Rightarrow Z \sim \text{Pois}(\lambda+\gamma)$

Imagine two ind. radioactive sources,