

If $X \sim \text{Bin}(n, p)$ then for large enough n , $X \dot{\sim} \mathcal{N}(np, np(1-p))$.

Moment Generating Function of $X \sim \text{Pois}(\lambda)$

PMF: $P_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$

MGF;

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tx}] = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{e^t \lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} \frac{e^{-\lambda} (e^t \lambda)^x e^{-e^t \lambda} e^{e^t \lambda}}{x!} \dots (1) \\ &= e^{e^t \lambda - \lambda} = e^{(e^t - 1)\lambda} \end{aligned}$$

where the above holds since $(1); \sum_{x=0}^{\infty} \frac{e^{-e^t \lambda} (e^t \lambda)^x}{x!} = 1$

So If $X_1, X_2, X_3, \dots, X_n$ are *iid* $\text{Pois}(1)$ then the Distribution of $X_1 + X_2 + \dots + X_n$ is ?

when Random Variables are independent;

$$\begin{aligned} M_{X_1+X_2+\dots+X_n}(t) &= M_{X_1}(t) M_{X_2}(t) \dots M_{X_n}(t) \\ &= (M_{X_1}(t))^n \\ &= \left(e^{(e^t-1)}\right)^n \end{aligned}$$

$$\text{so, } X_1 + X_2 + \dots + X_n \sim \text{Pois}(n)$$

and for n long enough; $X_1 + X_2 + \dots + X_n \dot{\sim} \mathcal{N}(n, n)$ So By Central Limit Theorem; If $X \sim \text{Pois}(n)$ then $X \dot{\sim} \mathcal{N}(n, n)$ for large enough n . {Use Continuity Correction}

→ Gamma convergence to Normal; (Exercise) (1) MGF of Gamma (n, λ) (2) Show that the MGF of $X_1 + X_2 + \dots + X_n$ is same as MGF of Gamma (n, λ) where X_1, X_2, \dots, X_n are *iid* $\text{Expo}(\lambda)$. (3) Then by CLT for n large enough; $X_1 + X_2 + \dots + X_n \dot{\sim} \mathcal{N}\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)$

Suppose Z_1, Z_2, \dots, Z_n are ind $\mathcal{N}(0, 1)$ then $\bar{Z}_n = \frac{Z_1 + Z_2 + \dots + Z_n}{n}$ will have $\bar{Z}_n \sim N\left(0, \frac{1}{n}\right)$. Note that \bar{Z}_n and $\sum_{j=1}^n (Z_j - \bar{Z}_n)^2$ are independent and $\sum_{j=1}^n (Z_j - \bar{Z}_n)^2 \sim \chi_{n-1}^2$

For $n = 2$; $\bar{Z}_n = \frac{Z_1 + Z_2}{2}$ and $\sum_{j=1}^2 (Z_j - \bar{Z}_2)^2 = (Z_1 - \bar{Z}_2)^2 + (Z_2 - \bar{Z}_2)^2$

$$\begin{aligned} &= \left(Z_1 - \frac{Z_1 + Z_2}{2} \right)^2 + \left(Z_2 - \frac{Z_1 + Z_2}{2} \right)^2 \\ &= \left(\frac{Z_1 - Z_2}{2} \right)^2 + \left(\frac{Z_2 - Z_1}{2} \right)^2 = 2 \left(\frac{Z_1 - Z_2}{2} \right)^2 = \left(\frac{Z_1 - Z_2}{\sqrt{2}} \right)^2 \end{aligned}$$

Note

$$\begin{aligned} Z_1 - Z_2 &\sim N(0, 2) \\ \frac{Z_1 - Z_2}{\sqrt{2}} &\sim N(0, 1) \\ \Rightarrow \left(\frac{Z_1 - Z_2}{\sqrt{2}} \right)^2 &\sim \chi_1^2(1) \end{aligned}$$

Now to show that \bar{Z}_2 and $\sum_{j=1}^2 (Z_j - Z_n)^2$ are independent, we need to show that $Z_1 + Z_2$ and $Z_1 - Z_2$, are independent since $\sum_{j=1}^2 (Z_j - Z_n)^2 = \left(\frac{Z_1 - Z_2}{\sqrt{2}} \right)^2$ is some function of Z_1 and Z_2 same as $\bar{Z}_2 = \frac{Z_1 + Z_2}{2}$ is a function of Z_1 and Z_2 .

Now If, $\bar{Z}_n \sim \mathcal{N}\left(0, \frac{1}{n}\right)$ then $\sqrt{n}\bar{Z}_n \sim \mathcal{N}(0, 1)$ and it is independent of $\sum_{j=1}^n (Z_j - Z_n)^2$ too.

Now, $T = \frac{Z}{\sqrt{\frac{X}{n}}}$ where $Z \sim \mathcal{N}(0, 1)$ and $X \sim \chi_n^2$ and,

$$\Rightarrow \frac{\sqrt{n}\bar{Z}_n}{\sqrt{\frac{\sum_{j=1}^n (Z_j - Z_n)^2}{n-1}}} \sim t_{n-1} \text{ distribution with } n-1 \text{ deg of freedom.}$$

Now suppose $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, then

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}$$

is the sample variance, when \bar{X}_n is the sample mean, and the distribution of $\frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$.

Note that,

$$\begin{aligned}\frac{(n-1)S_n^2}{\sigma^2} &= \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \\ &= \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} - \frac{\bar{X}_n - \mu}{\sigma} \right)^2 \\ &= \sum_{i=1}^n (Z_i - \bar{Z}_n)^2 \sim \chi_{n-1}^2\end{aligned}$$

and

$$\begin{aligned}Z_i &= \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, 1) \\ \bar{Z}_n &= \frac{\sum \frac{X_i - \mu}{\sigma}}{\sigma} = \frac{\sum (X_i - \mu)}{n\sigma} = \bar{Z}_n\end{aligned}$$

$$\begin{aligned}\bar{Z}_n &= \frac{\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)}{n} \text{ and } \frac{(n-1)s_n^2}{\sigma^2} \\ &\text{are independent by } \dots (1) \\ \bar{Z}_n &= \frac{\bar{X}_n - \mu}{\sigma} \sim N\left(0, \frac{1}{n}\right) \\ \text{then } \sqrt{n}\bar{Z}_n &= \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \sim \mathcal{N}(0, 1) \\ \Rightarrow \frac{\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right)}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2(n-1)}}} &= \frac{\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right)}{\sqrt{\frac{(S_n)^2}{\sigma^2}}} \sim t_{n-1}\end{aligned}$$

Now, $\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \sim N(0, 1)$ in this we can replace σ by S_n which will then give us t_{n-1} distribution and will have a fatter tail than $\mathcal{N}(0, 1)$.

Statistics

Point Estimation

When we are trying to estimate a finite number of points or parameter; e.g; trying to estimate the average height of the class of some students, where average height = μ So population μ is the unknown parameter of interest. Say distribution of the whole Class's average height is $\mathcal{N}(\mu, 1)$ then we draw a sample $X_1, X_2, X_3, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, 1)$ then we find a reasonable function $h(X_1, X_2, \dots, X_n)$ which will be equal to the estimator of μ .

one good estimator of μ is $\frac{X_1+X_2+\dots+X_n}{n}$, the sample mean, and it is good estimator because by law of large numbers as $n \rightarrow \infty$ $\frac{X_1+X_2+\dots+X_n}{n}$ converges to μ in probability.

Biased and Unbiased Estimators

(1) $\hat{H}_1 = X_1$ (is unbiased but not a very good estimator, since variance does not tend to 0)

(2) $\hat{H}_2 = \frac{\sum_{i=1}^n X_i}{n}$ (is unbiased and also a good estimator)

(3) $\hat{H}_3 = \frac{\sum_{i=1}^n X_i}{n} + \frac{1}{n}$ is a biased estimator.