Normal Dist (Cont.)

$Z \sim N(0,1)$, CDF Φ, $\mathbb{E}(Z) = 0$, $\text{Var}(Z) = 1$

$-Z \sim N(0,1)$ By Symmetry ($e^{-z^2/2}$, Bell Curve Symmetric)

Let $X = \mu + \sigma Z$, $\mu \in \mathbb{R}$ (mean, location), $\sigma > 0$ (SD, Scale)

then $X \sim N(\mu, \sigma^2)$. Since

$\mathbb{E}(\mu + \sigma Z) = \mu$ $\quad\quad$ $\text{Var}(\mu + \sigma Z) = \sigma^2 \, \text{Var}(Z) = \sigma^2$

↳ Simple Shift

Proof.

in general $\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$

$\Rightarrow$ $\text{Var}(X + c) = \text{Var}(X) = \mathbb{E}\left((X + c - \mathbb{E}(X+c))^2\right) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$

$\text{Var}(cX) = c^2 \cdot \text{Var}(X) = \mathbb{E}\left((cX - \mathbb{E}(cX))^2\right) = \mathbb{E}\left(c^2 \cdot (X - \mathbb{E}(X))^2\right)$

Sidenote $\Rightarrow$ $\text{Var}(X) \geq 0$; $\text{Var}(X) = 0$ iff $P(X = a) = 1$ for some a.

$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$ in general.

[They're equal if X and Y are indep.]

$\text{Var}(X + X) \neq \text{Var}(X) + \text{Var}(X)$ since X is NOT iid with itself

$Var(X+X) \neq Var(X) + Var(X)$ since X is NOT i.i.d with itself

X is actually fully dependent with itself. (Not the same as $X_1 + X_2$)

$Var(X+X) = Var(2X) = 4 \cdot Var(X)$

Also, $Z = \frac{X-\mu}{\sigma}$ (standardization)

$\Rightarrow$ Variance σ^2 is always positive

PDF of $X \sim N(\mu, \sigma^2)$

CDF: $P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right)$

$= \Phi\left(\frac{x-\mu}{\sigma}\right)$ taking the Derivative

$\Rightarrow PDF(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2}$

e.g. $-X + \sigma(-Z). \sim N(-\mu, \sigma^2)$

Later we'll show,

If $X_j \sim N(\mu_j, \sigma_j^2)$ indep, then

$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

68- 95- 99.7 Rule : $X \sim N(\mu, \sigma^2)$

$$P(|X-\mu| \leq \sigma) \approx 0.68$$

$$P(|X-\mu| \leq 2\sigma) \approx 0.95$$

$$P(|X-\mu| \leq 3\sigma) \approx 0.997$$

Some More Variances (with the help of Lotus)

Reminder : LOTUS : $\mathbb{E}(g(X)) = \sum_x g(x) P(X=x)$

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

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

Strategy : Starting from $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$

$\frac{d}{d\lambda} \implies \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{k!} = e^{\lambda}$ $\left[= \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda} \right]$

$\cdot \lambda \implies \sum_{k=1}^{\infty} \frac{k\lambda^k}{k!} = \lambda e^{\lambda}$

$\cdot \lambda \implies \sum_{k\geq 1} \frac{k\lambda}{k!} = \lambda e^{\cdot\cdot}$

$\frac{d}{d\lambda} \implies \sum_{k\geq 1}^{\infty} \frac{k^2 \lambda^{k-1}}{k!} = e^{\lambda} + \lambda e^{\lambda} \xrightarrow{\cdot \lambda} \sum_{k\geq 1}^{\infty} \frac{k^2 \lambda^k}{k!} = \lambda(1+\lambda)e^{\lambda}$

$\implies \mathbb{E}(x^2) = e^{-\lambda} \times \lambda \times (1+\lambda) e^{\lambda} = \lambda^2 + \lambda$

$Var(x) = \mathbb{E}(x^2) - \underbrace{\mathbb{E}^2(x)}_{} = \lambda^2 + \lambda - \lambda^2 = \lambda$

$\mathbb{E}(x)$ Proven to be λ before.

$\implies X \sim Pois(\lambda)$ has both mean and Variance λ!

Consider $X \sim Bin(n, p)$:

$X = I_1 + \dots + I_n \; ; \; I_j \overset{iid}{\sim} Bern(p)$

$x^2 = I_1^2 + \dots + I_n^2 + 2I_1 I_2 + \dots + 2I_{n-1} I_n$

Linearity of expectation   Binomial # Pairs

$\mathbb{E}(x^2) = n \cdot \mathbb{E}(I_1^2) + 2\binom{n}{2}\mathbb{E}(I_1 I_2)$

$I_i = \{0, 1\}$

$= n \cdot \mathbb{E}(I_1) + n(n-1) \cdot P($ Both trials Succeeding $)$

$$= n \cdot p + n(n-1) \cdot p^2 = np + n^2p^2 - np^2$$

Therefore

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = np + n^2p^2 - np^2 - (np)^2 = np - np^2$$

$$= np(1-p) = npq$$

$$Var(X) = npq$$

Proof of LOTUS for discrete Sample Space:

Imagine Pebble world again

$$\underbrace{\sum_x g(x) P(X=x)}_{\text{grouped}} = \sum_{s \in S} \underbrace{g(X(s)) P(\{s\})}_{\text{ungrouped}}$$

$$= \sum_x \sum_{s: X(s)=x} g(X(s)) P(\{s\}) = \sum_x g(x) \sum_{s: X(s)=x} P(\{s\})$$

Summing over ungrouped possible values

$$= \sum_x g(x) P(X=x)$$