Probability: Sums of Random Variables

Variance of Sums of Random Variables

- Covariance
- Correlation

Normal Distribution

Central Limit Theorem

Examples

Variance of Sums

$$Var(X) = E(X-\mu)^2 = E(X^2) - (EX)^2$$

$$Var(X+Y) = E[(X+Y)^2] - [EX + EY]^2$$

$$= E[X^2 + 2XY + Y^2] - [(EX)^2 + 2EXEY + (EY)^2]$$

$$= E(X^2) - (EX)^2 + 2(EXY - EXEY) + E(Y^2) - (EY)^2$$

$$= Var X + 2\,Cov(X,Y) + Var Y$$

$$Cov(X,Y) = EXY - EXEY$$

Note! If X and Y are independent,

$$Cov(X,Y) = 0$$

and $Var(X+Y) = Var X + Var Y$

Properties of Covariance

$$(aX+bY)(cZ+dW) = acXZ + adXW + bcYZ + bdYW$$

$$\mathrm{Cov}(aX+bY,\ cZ+dW)$$
$$= ac\,\mathrm{Cov}(X,Z) + ad\,\mathrm{Cov}(X,W) +$$
$$+\ bc\,\mathrm{Cov}(Y,Z) + bd\,\mathrm{Cov}(Y,W)$$

$$\mathrm{Cov}(X,X) = E(XX) - EXEX$$
$$= E(X^2) - (EX)^2 = \mathrm{Var}\,X$$

$$\mathrm{Var}(aX) = \mathrm{Cov}(aX, aX) = a^2\,\mathrm{Var}\,X$$

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X}\,\sqrt{\mathrm{Var}Y}}$$

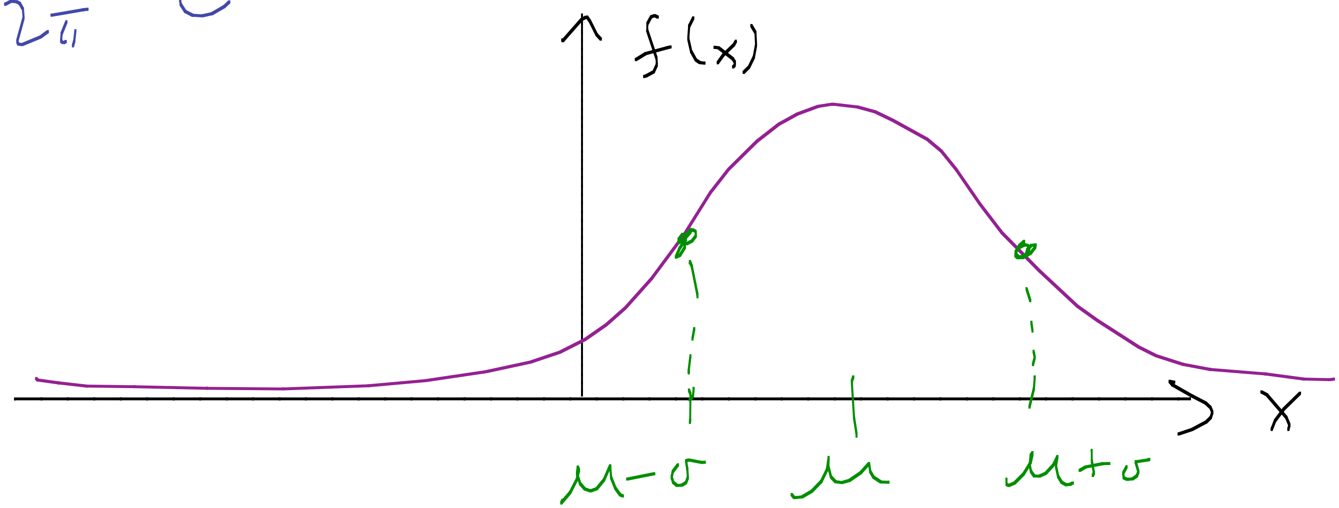
$$-1 \le \mathrm{Corr}(X,Y) \le 1$$

Normal distribution aka Gaussian

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

mean variance

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$F(x)$ – has no elementary formula

$\Phi(z) = P[Z \leq z]$ $Z \sim N(0,1)$

When using the normal distribution, choose the nearest z-value to find the probability, or if the probability is given, choose the nearest z-value. No interpolation should be used.

Example: If the given z-value is 0.759, and you need to find $\Pr(Z < 0.759)$ from the normal distribution table, then choose the probability for z-value = 0.76: $\Pr(Z < 0.76) = 0.7764$.

When using the normal approximation to a discrete distribution, use the continuity correction.

$$\Phi(-.3) = P[Z \le -.3)$$

$$= P[Z \ge .3] = P[Z > .3]$$

$$= 1 - P[Z \le .3)$$

$$\Phi(-.3) = 1 - \Phi(.3)$$

$$\Phi(-z) = 1 - \Phi(z)$$

Normal Distributions and Sums

$X, Y \sim$ jointly normal

$$X \sim N(\mu_X, \sigma_X^2), \quad Y \sim N(\mu_Y, \sigma_Y^2)$$

Then $X+Y$ is $N\left[\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\,\mathrm{Cov}(X,Y)\right]$

Central Limit Theorem

$X_1, X_2, \ldots$ iid random variables

$$S_k = X_1 + \cdots + X_k$$

Then
$$\frac{S_k - ES_k}{\sqrt{\mathrm{Var}\, S_k}} \xrightarrow[\infty]{k} N(0,1)$$

Normal Approximation

$$\frac{S_k - ES_k}{\sqrt{\operatorname{Var} S_k}} \approx Z \sim N(0,1)$$

$$P[S_k \le x] = P\left[\frac{S_k - ES_k}{\sqrt{\operatorname{Var} S_k}} \le \frac{x - ES_k}{\sqrt{\operatorname{Var} S_k}}\right]$$

$$\approx P\left[Z \le \underbrace{\frac{x - ES_k}{\sqrt{\operatorname{Var} S_k}}}_{z\text{-value}}\right]$$

Continuity Correction

Suppose S_k is integer valued, $x \in \mathbb{Z}$

$$P[S_k = x] > 0$$

but $P[\text{normal} = x] = 0$ b/c normals are continuous

If $S_k \approx Y$ $\quad Y \sim N(ES_k, \operatorname{Var} S_k)$

then $P[S_k = x] \approx P[x - \frac{1}{2} < Y \leq x + \frac{1}{2}]$

$$P[S_k \leq x] \approx P[S_k \leq x + \tfrac{1}{2}]$$

$$\approx P\left[Z \leq \frac{x + \frac{1}{2} - ES_k}{\sqrt{\operatorname{Var} S_k}}\right]$$

The average height of adult Americans is 176 cm, with a standard deviation of 6 cm, for males, and 163 cm, with a standard deviation of 5 cm, for females. If heights of each group are normally distributed, what is the probability that a randomly selected $\underbrace{\text{American male}}_{M}$ is taller than a randomly selected $\underbrace{\text{American female}}_{F}$?

A. 0.85 B. 0.88 C. 0.91 D. 0.93 $\boxed{\text{E. } 0.95}$

$$M \sim N(176, 36)$$

$$F \sim N(163, 25)$$

$$M - F \sim N(13, 36 + (-1)^2 \cdot 25) = N(13, 61)$$

$$P[M > F] = P\left[\frac{M-F-13}{\sqrt{61}} > \frac{0-13}{\sqrt{61}}\right]$$

$$= 1 - \Phi\left(\frac{-13}{\sqrt{61}}\right) = \Phi\left(\frac{13}{\sqrt{61}}\right) = \Phi(1.66) \approx .95$$

Suppose that X and Y are jointly normal random variables, with $\mathrm{E}X = 1$, $\mathrm{Var}X = 4$, and $\mathrm{E}Y = -2$, $E\left(Y^2\right) = 5$. If the correlation of X and Y is $-1/2$, what is the probability that the sum of X and Y is positive?

A. .16 B. $\boxed{.28}$ C. .37 D. .72 E. .84

$X + Y \sim N(-1, 3)$

$E(X+Y) = EX + EY = 1 - 2 = -1$

$\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y + 2\,\mathrm{Cov}(X, Y)$

$= 4 + [5 - (-2)^2] + 2\,\mathrm{Corr}(X,Y)\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}$

$= 4 + 1 + 2\left(-\frac{1}{2}\right)\cdot 2 \cdot 1 = 3$

$P[X+Y>0] = P\left[\frac{X+Y-{-1}}{\sqrt{3}} > \frac{0-{-1}}{\sqrt{3}}\right] = 1 - \Phi(.58)$

$= 1 - .72 = .28$

If I flip a fair coin 100 times, what is the approximate probability that I will get at least 45 heads?

$X_1 = 1$ if 1^{st} flip is heads

$\vdots$

$X_k = 1$ if k^{th} flip is heads

A. Less than 0.80

B. At least 0.80, but less than 0.82

C. At least 0.82, but less than 0.84

D. At least 0.84, but less than 0.86

E. At least 0.86

$$S_{100} = \sum_{k=1}^{100} X_k$$

S_{100} is discrete, so need continuity correction.

$$E\,S_{100} = \frac{1}{2} \cdot 100 = 50, \quad \mathrm{Var}\,X_1 = E(X_1^2) - (E X_1)^2$$

$$= \frac{1}{2} \cdot 1 - \frac{1}{4} = \frac{1}{4}$$

$$\mathrm{Var}\,S_{100} = 100 \cdot \frac{1}{4} = 25$$

$$= npq = 100\left(\frac{1}{2}\right)\left(1 - \frac{1}{2}\right)$$

If I flip a fair coin 100 times, what is the approximate probability that I will get at least 45 heads?

$ES_H = 50$

$\mathrm{Var}\, S_H = 25$

A. Less than 0.80
B. At least 0.80, but less than 0.82
C. At least 0.82, but less than 0.84
D. At least 0.84, but less than 0.86
$\boxed{\text{E. At least 0.86}}$

$$P[S_H \geq 45] = P[S_H \geq 44.5]$$

$$\approx P\left[Z \geq \frac{44.5 - 50}{\sqrt{25}}\right]$$

$$= 1 - \Phi\left(-\frac{5.5}{5}\right) = \Phi(1.1) \approx .8643$$