Random Variables Related

I Sum of two random variables

$X \sim$ densities $f(x)$ ⎫ two independent
$Y \sim$ density $g(x)$ ⎭ random variables

$$P(X+Y \leq z) = \int P(X+Y \leq z \mid Y=y)\, g(y)\, dy$$

$$= \int P(X \leq z-y)\, g(y)\, dy$$

→ Take derivative w.r.t z

→ $h(z) = \int_{-\infty}^{\infty} f(z-y)\, g(y)\, dy$

$$\boxed{f_{X+Y}(z) = \int_{\mathbb{R}} f(x, z-x)\, dx}$$

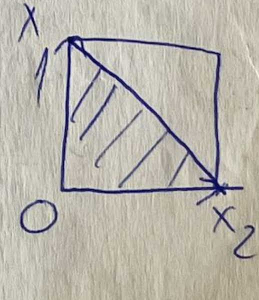
→ change $\int_{-\infty}^{\infty}$ to y.

II Sum of n random variables

$X_1, \dots, X_n$ i.i.d, $U[0,1]$

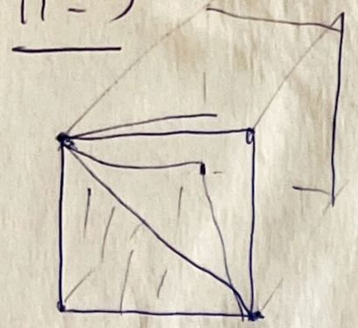
⊕ $P(X_1 + X_2 + \dots + X_n \leq 1)$

strat with $\underline{n=2}$:



$$P(X_1 + X_2 \leq 1) = \frac{1}{2}$$

For $\underline{n=3}$



$$P(X_1+X_2+X_3 \leq 1)$$
$$= V = \frac{1}{3} A \cdot h = \frac{1}{3} \times \frac{1}{2} \times 1$$
$$= \frac{1}{6}$$

→ suggest: $P(X_1+X_2+\dots+X_n \leq 1)$
$$= \frac{1}{n!}$$

→ prove by induction:
start with n and deduce $n+1$

→ Need to prove: $P(X_1+X_2+\dots+X_{n+1} \leq 1) = \frac{1}{(n+1)!}$

$$P(X_1+X_2+\dots+X_{n+1} \leq 1) = \int_0^1 P(X_1+X_2+\dots+X_n \leq 1-X_{n+1}) \underbrace{f(X_{n+1})}_{=1} \, dX_{n+1}$$

since X_is are uniform

→ Need:

$$P(X_1+X_2+\dots+X_n \leq 1-X_{n+1})$$

→ Need to shrink every dimension of the n-dimension simplex from 1 to $(1-X_{n+1})$ → $\frac{(1-X_{n+1})^n}{n!}$

(again start from $n=2$)

III Product of Random variables:

$$Z = X \cdot Y$$

- Proceed similar to sum
- $cov(X,Y) = 0$, are X, Y independence?

$\Rightarrow$ use $Y = ZX$, Z, X are independence. $P\begin{pmatrix} Z=-1 \\ Z=1 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

$$\begin{aligned} cov(X,Y) &= cov(X, ZX) \\ &= E(X^2 Z X) - E(X)E(Z)E(X) \\ &= 0 \end{aligned}$$

But X, Y are not independence.

IV Stick Breaking

(+) $X, Y \sim U[0,1]$ ~ separate $[0,1]$ into three pieces. what is the probability that the 3 resulting pieces form a triangle?

- A triangle can only be formed if the largest pieces is less than the sum of the other two.
- assuming $X > Y$ or $X \le Y$

$\Rightarrow$ Break into $X, Y-X, 1-Y$

⊕ $X, Y, Z \in U[0,1]$

- suppose Z is fixed and be the longest element.
- The area is bounded by:

$$0 \le X \le Z \qquad \cancel{0}\, Z \le X+Y$$
$$0 \le Y \le Z$$
$$\cancel{0 \le Z \le}$$

→ 

$$\rightarrow \text{volume} = \int_0^1 \frac{Z^2}{2}\, dZ = \frac{1}{6}$$

- Any element can be the largest: $\frac{1}{6} \times 3 = \frac{1}{2}$

V order statistics

1) Max and min

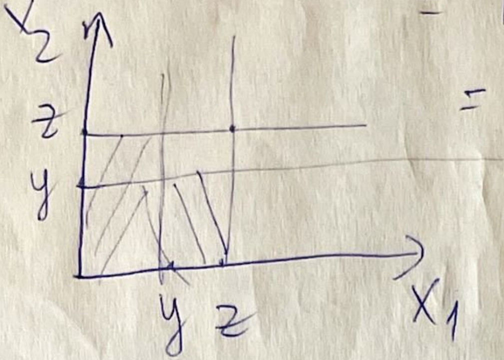
$$Y_n = \min(X_1, \dots, X_n) \qquad X_i \text{ - i.i.d}$$
$$Z_n = \max(X_1, \dots, X_n)$$
$$P(Y_n \ge x) = \left(P(X_n \ge x)\right)^n = (1 - F_X(x))^n$$
$$P(Z_n \le x) = (F_X(x))^n$$

⊕ Correlation of max, min

$$\operatorname{corr}(Y,Z) = \frac{\operatorname{cov}(Y,Z)}{\sqrt{\operatorname{var}(Y)}\sqrt{\operatorname{var}(Z)}} = \frac{E(YZ) - \overbrace{E(Y)E(Z)}^{\text{Easy}}}{\underbrace{\sqrt{E(Y^2)-E(Y)}\sqrt{E(Z^2)-E(Z)}}_{\text{Easy to compute}}}$$

→ need to calculate $E[YZ]$

→ Need $F(YZ) = P(Y \le y, Z \le z)$



$= P(Z \le z) - P(Y \ge y \cap Z \le z)$

$= z^2 - (z-y)^2$

$= 2zy - y^2$

→ $f(YZ) = \dfrac{\partial F(YZ)}{\partial Y \partial Z} = 2$

→ $E(YZ) = \int_0^1 \int_0^z 2yz \, dy \, dz$

⊕ Muti-variate normal:

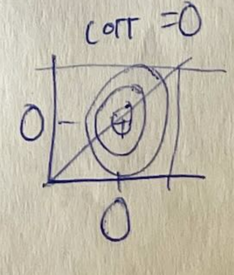
$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & p \\ p & 1 \end{bmatrix} \right)$$

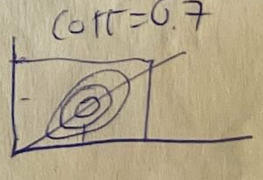
$E(\max(X, Y))$, Let $Z = \max(X, Y)$, ~~assum~~

$E(Z) = E(Z \mid X \le Y) P(X \le Y) + P(Z \mid X \ge Y) P(X \ge Y)$

$= 2E(Z \mid X \ge Y) P(X \ge Y)$

$\boxed{P(X \ge Y) = \dfrac{1}{2}}$





$$E(Z) = E(X \mid X > Y)$$

$$X = W_1$$

$$\cancel{Y =}\ W_2 = X - Y$$

$$= E(W_1 \mid W_2 > 0)$$

$$= E(E(W_1 \mid W_2) \mid W_2 > 0)$$

k^{th} order statistics:

- k^{th} smallest value
- $P(X_{(k)} \leq x)$: need at least k of the n X_i's less than or equal to x.

$\Rightarrow P(X_{(k)} \leq x) = P(\text{at least } k \ X_i \leq x)$

$$= \sum_{j=k}^{n} \binom{n}{j} (F(x))^j (1-F(x))^{n-j}$$

density function:

$$\frac{d}{dx} P(X_{(k)} \leq x)$$

$$= \frac{d}{dx} \left(\sum_{j=k}^{n} \binom{n}{j} (F(x)^j)(1-F(x))^{n-j} \right.$$

$$= \sum_{j=k}^{n} \binom{n}{j} \Big(jF(x)^{j-1} f(x) (1-F(x))^{n-j}$$

$$- (n-j)(1-F(x))^{n-j-1} \cdot f(x) \Big)$$

$$= n f(x) \sum_{j=k}^{n-1} \binom{n-1}{j-1} \Big(F(x)^{j-1} (1-F(x))^{n-j}$$

$$- \binom{n-1}{j} (F(x)^j)(1-F(x))^{n-j-1} \Big)$$

$$+ n f(x) F(x)^{n-1}$$

VI Convolution:

⊕ $X : f_X(x) \rightarrow$ density $g(X)$

$$Y = g(x)$$

$$P(Y \leq y) = P(g(x) \leq y) = P(x \leq g^{-1}(y))$$
$$= F_X(g^{-1}(y))$$

$$\Rightarrow f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y))$$
$$= \frac{d}{dy}(g^{-1}(y)) f_X(g^{-1}(y))$$

$$\frac{d}{dy}(g^{-1}(y))$$

$$y = y$$
$$g(g^{-1}(y)) = y$$
$$\rightarrow \frac{d}{dy} g(g^{-1}(y)) = 1$$
$$\rightarrow g'(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = 1$$
$$\rightarrow \frac{d}{dy} g^{-1}(y) = \frac{1}{g'(g^{-1}(y))}$$

⊕ Is $E(Y) = g(E(X))$?

Jensen's inequality - for convex function

$g(x) = x^2$ - example, $\boxed{E(g(x)) \geq g(E(X))}$

$\Leftrightarrow E(X^2) \geq E(X)^2 \Leftrightarrow var(X) \geq 0$

VII <u>Moments</u>

: 1st: Mean , 2nd: Variance

3rd Moments: skewness

4th Moments: Kurtosis

$$\phi_n = E\left(\left(\frac{X-E(X)}{S(X)}\right)^n\right]$$

$$\phi_n^+ = E\left(\left[\frac{X-E(X)}{S(X)}\right]^n X > E(X)\right] P(X > E(X))$$

θ_n^-

<u>for n odds</u>: $\phi_n = \phi_n^+ - \phi_n^-$ measures

if skewness measure the asymmetry of the probability distribution

<u>for n even</u> $\phi_n = \phi_n^+ + \phi_n^-$ - higher value

if x far from $E(X)$

→ measure <u>how fat the</u> <u>tails are</u>

$M_X(t) = E(e^{tX})$ → Normal: $\int e^{tx} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} dx$

$= e$

$M(X^n) = M_X^{(n)}(0)$

$\underline{\text{VIII}}$ sampling related:

$$\bar{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$$ – Normal distribution

$$\bar{\sigma}_i^2 = \frac{1}{n-1}\sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2$$ – chi-square distribution

~~histogram,~~ | ~~distribution of $\bar{y}_i$~~ $\bar{u}_i$ will be normal

sq sum of square normal distribution

standard

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$$

$$\rightarrow \frac{s_0^2}{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} z_i^2 \sim \frac{1}{n}\chi_n^2$$

sample mean is used instead of the population ~~sur~~ mean.

$$\rightarrow \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 \sim \chi_{n-1}$$

Correlation problems:

1. Constructing correlated random variables

⊕ $X \sim N(0, a)$ with $\text{corr}(X, Y) = \rho$

$Y \sim N(0, b)$

let $Z_1 \sim N(0,1)$ $Z_2 \sim N(0,1)$ independent

set $X = a Z_1$

$Y = b(\rho Z_1 + \sqrt{1-\rho^2}\, Z_2)$

2. optimal hedge ratio

① $Var(A) = \sigma_A^2$

$Var(B) = \sigma_B^2$

portfolio $A - hB$

$$\underset{h}{\operatorname{argmin}} \; Var(A - hB)$$

$$= \sigma_A^2 + h^2\sigma_B^2 - 2\rho h \sigma_A \sigma_B$$

$$\frac{\partial}{\partial h} = 2h\sigma_B^2 - 2\rho\sigma_A\sigma_B$$

$$\Rightarrow h^* = \rho\frac{\sigma_A}{\sigma_B}$$

$$\frac{\partial^2}{\partial h^2} = 2\sigma_B^2 > 0 \rightarrow h^* \text{ is minimum}$$

② Mean - variance optimization

- starting W_{t-1} → choosing $x_{t-1} = (x_{t-1}^1, \ldots, x_{t-1}^N)$ to invest in N assets

- $R_t = (R_t^1, \ldots, R_t^N) = R^f + R_t^e$

$$\rightarrow W_t^{x_{t-1}} = W_{t-1}(1+R^f) + x_{t-1}^T R_t^e$$

pwc

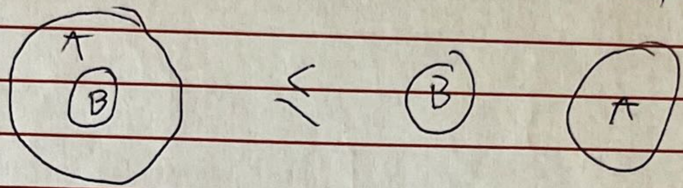
3. Correlation between variables

(+) Two variables

$P(A) = x \qquad P(B) = y$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$

$\geq \max(P(A), P(B))$ · A occurs, B do not
B occurs, A do not



$P(A \cap B) = P(A)P(B) - \mathrm{cov}(A, B)$

$= P(A)P(B) - \rho_{AB} \sqrt{\mathrm{Var}(A)\,\mathrm{Var}(B)}$

→ Use $P(A \cup B)$ to find the range of ρ_{AB}

(+) Three variables

The covariance matrix must be p.d

⟹ det of small matrix is positive

~~$\mathrm{cov}(a^T X) = a^T \mathrm{cov}(X) a$~~

$Y = a^T X \rightarrow \mathrm{Var}(Y) = a^T \mathrm{cov}(X) a \geq 0$