Independence and Expectation:

$(X_1, X_2, \dots, X_n) \sim f(x_1, \dots, x_n)$.

Suppose $X_1, \dots, X_n$ are mutually independent.

Then $f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

$$\underline{\mathbb{E}[X_1 X_2 \dots X_n]}$$

$$= \int (x_1 x_2 \dots x_n) f(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

$$= \int \underline{(x_1 f_{X_1}(x_1))}\, \underline{(x_2 f_{X_2}(x_2))} \dots \underline{(x_n f_{X_n}(x_n))}\, dx_1 \dots dx_n$$

$$= \left(\int x_1 f_{X_1}(x_1)\, dx_1\right) \dots \left(\int x_n f_{X_n}(x_n)\, dx_n\right)$$

$$= \underline{\mathbb{E}X_1 \cdot \mathbb{E}X_2 \dots \mathbb{E}X_n}$$

$$\Rightarrow \operatorname{Cov}(X_i, X_j) = 0, \; i \neq j$$

If $X_1, \ldots, X_n$ are mutually indep. random variables, then $g_1(X_1), g_2(X_2), \ldots g_n(X_n)$ are also independent.

Look at

$$P[g_1(X_1) \in A_1, g_2(X_2) \in A_2, \ldots, g_n(X_n) \in A_n]$$

$$= P[X_1 \in g_1^{-1}(A_1), X_2 \in g_2^{-1}(A_2), \ldots, X_n \in g_n^{-1}(A_n)]$$

$$= \prod_{j=1}^{n} P(X_j \in g_j^{-1}(A_j))$$

$$= \prod_{j=1}^{n} P(g_j(X_j) \in A_j)$$

Hence, independent.

Likewise: $g_1(X_1, \ldots, X_{i_1}), g_2(X_{i_1+1}, \ldots, X_{i_2}), g_3(\ldots), \ldots$ and so on.. are independent.]

Note: $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

Immediate from the definition of covariance!

Look at a linear combination of a generic r. vector $(X_1, \ldots, X_n)$:

$$V = a + b_1 X_1 + \cdots + b_n X_n.$$

$$\mathbb{E}V = a + b_1 \mathbb{E}X_1 + \cdots + b_n \mathbb{E}X_n$$

$$\mathrm{Var}(V) = \mathrm{Var}(a + b_1 X_1 + \cdots + b_n X_n)$$

$$= \mathrm{Var}(b_1 X_1 + \cdots + b_n X_n)$$

$$= \mathrm{Cov}(b_1 X_1 + \cdots + b_n X_n,\ b_1 X_1 + \cdots + b_n X_n)$$

Bilinear operator

Verify

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} b_i^2 \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Var}(b_1 X_1 + \dots + b_n X_n)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \,\mathrm{Cov}(X_i, X_j)$$

$$= \underset{\sim}{b}^T \Sigma \underset{\sim}{b} \geq 0$$

$$\Sigma = \mathrm{Disp}(\underset{\sim}{X}). \quad \underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$[\mathrm{Disp}(\underset{\sim}{X})]_{i,j} = \mathrm{Cov}(X_i, X_j).$$

symmetric

Σ must necessarily be a non-negative definite matrix.

Very useful results

$$Y_{m\times 1} = A_{m\times n} X_{n\times 1}$$

$$\mathbb{E}Y = A\,\mathbb{E}X.$$

$$\mathrm{Disp}(Y) = A \Sigma A^T.$$

Moment-generating functions of random vectors:

$$\underset{\sim}{X}_{n\times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = \mathbb{E}\left[e^{\underset{\sim}{t}^T \underset{\sim}{X}}\right] \qquad \underset{\sim}{t} \in \mathbb{R}^n$$

$$= \mathbb{E}\left[e^{\sum t_i X_i}\right]$$

is said to exist if it's finite for all $\|\underset{\sim}{t}\| < \varepsilon_0$ for some $\varepsilon_0 > 0$.

<u>FACTS</u>: (a) If $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are two n dim random vectors with $M_{\underset{\sim}{X}}$ and $M_{\underset{\sim}{Y}}$ existing finitely for all sufficiently small $\underset{\sim}{t}$ (i.e all $\|\underset{\sim}{t}\| < \varepsilon_0$ for some ε_0) and $M_{\underset{\sim}{X}}(\underset{\sim}{t}) = M_{\underset{\sim}{Y}}(\underset{\sim}{t})$ for all such $\underset{\sim}{t}$, then $X \stackrel{d}{=} Y$.

(b) $\{\underline{X}_n\}$ of random vectors and $\{\underline{X}_n\}$ and $\underline{X}$ all have finite mgf's for all small $\underset{\sim}{t}$, then

$\underline{X}_n \xrightarrow{d} \underline{X}$ if and only if

$[M_{\underline{X}_n}(\underset{\sim}{t}) \longrightarrow M_{\underline{X}}(\underset{\sim}{t})$ for all such $\underset{\sim}{t}$.

As in the 1 dimensional case, convergence in distribution is defined as $\mathbb{F}_{\underline{X}_n}(\underline{x_1, x_2, \ldots, x_n})$

$\longrightarrow \mathbb{F}_{\underline{X}}(\underline{x_1, \ldots, x_n})$ at every continuity point $(x_1, \ldots, x_n)$ of $\mathbb{F}_{\underline{X}}$.

Let's compute mgf's of some random vectors.

$(X_1, X_2, \ldots, X_n)$ are i.i.d $Ber(p)$.

Joint p.m.f of $(X_1, \ldots, X_n)$.

$$f_{X_1,\ldots,X_n}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$$

$$= P(X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots, X_n = \varepsilon_n)$$

$$= \prod_{i=1}^{n} P(X_i = \varepsilon_i) = p^{\sum \varepsilon_i}(1-p)^{n-\sum \varepsilon_i}$$

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$M_{S_n}(t) = \mathbb{E}[e^{tS_n}]$$

$S_n = X_1 + \ldots + X_n$

$$= \mathbb{E}\left[e^{t(X_1 + \ldots + X_n)}\right]$$

$$= \mathbb{E}\left[e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_n}\right]$$

are all independent

$$= \prod_{j=1}^{n} \mathbb{E}[e^{tX_j}] \qquad X_j \sim Ber(p)$$

So $M_{S_n}(t) = \left[\mathbb{E}(e^{tX_1})\right]^n$

$$\mathbb{E}\left[e^{tX_1}\right] = e^t p + (1-p)$$

so: $M_{S_n}(t) = \left(pe^t + (1-p)\right)^n$.

Exercise: Show using mgf's ~~that~~ if $X_1, \dots, X_n$ are independent and $X_i \sim Poi(\lambda_i)$, then $S_n = X_1 + \dots + X_n \sim Poi(\lambda_1 + \dots + \lambda_n)$.

Compute $\mathbb{E}\left[e^{tS_n}\right]$ and show it is the mgf of $Poi(\lambda_1 + \dots + \lambda_n)$.

Another related problem:

Suppose $Y_1, \dots, Y_r$ are i.i.d Geometric p random variables where Y_i counts ~~the~~ number of trials needed to get the first H.

$S_r = Y_1 + Y_2 + \dots + Y_r$.

What is the distribution of S_r?

Guesses? S_r is $NB(r, p)$

counts ~~the~~ number of trials to get r Heads.

r people doing Geometric.

Person 1. $\underbrace{TTTT\dots H}_{m_1 \text{ trials}} \rightarrow Y_1 = m_1$

Person 2 $\underbrace{TT\dots TH}_{m_2 \text{ trials}}$ Y_2

$r=2$

concatenate these sequences.

$\boxed{TT\dots THT\dots TH}$

this is exactly # of trials needed for 2H!

$r = 2$

$$TT \cdots TH \quad T \cdots TH$$

mgf of S_r

$$\mathbb{E}\left[e^{t(Y_1 + Y_2 + \cdots + Y_r)}\right]$$

$$= \prod_{i=1}^{r} \mathbb{E}\left[e^{tY_i}\right]$$

$$= \mathbb{E}\left[e^{tY_1}\right]^{r}.$$

$$P(Y_1 = j) = (1-p)^{j-1} p, \quad j \geq 1$$

So: $$\mathbb{E}\left[e^{tY_1}\right] = \sum_{j=1}^{\infty} e^{tj}(1-p)^{j-1} p$$

$$= \sum_{j=1}^{\infty} (e^t p)\left((1-p)e^t\right)^{j-1}$$

$$= e^t p \sum_{j=1}^{\infty} \left[\underbrace{(1-p)}_{1-p^*} e^t\right]^{j-1} \leftarrow \left[1 - \underbrace{(1-p)e^t}_{p^*}\right]$$

$$\frac{e^t p}{p_*} \left\{ \boxed{\sum_{j=1}^{\infty} (1-p_*)^{j-1} p_*} \right.$$

where $p_* = 1 - (1-p)e^t$

$$\left. = \frac{e^t p}{1-(1-p)e^t} \right\}.$$

$$\mathbb{E}[e^{tS_r}] = \left(\frac{e^t p}{1-(1-p)e^t} \right)^r .$$

Now need to compute: $\underline{W \geq r}$

$\mathbb{E}[e^{tW}]$ directly. $\underline{\underline{W}} \sim NB(r,p)$

$$\underline{P[W = m]} = \binom{m-1}{r-1} \underline{p^r (1-p)^{m-r}}$$

$$\mathbb{E}[e^{tW}] = \mathbb{E}[e^{t(W-r)+tr}]$$

$$= e^{tr}\, \mathbb{E}[e^{t\tilde{W}}]$$

$\underline{\underline{\tilde{W}}} = \underline{\underline{W-r}}$ = # of T's before r'th H

From $\tilde{W}$,

$$\mathbb{E}\left[e^{tS_r}\right] = e^{tr}\left[\frac{p}{1-(1-p)e^{t}}\right]^{r}$$

you should be able to show by direct calculation ~~that~~ this is $\mathbb{E}\left[e^{t\tilde{W}}\right]$

$$P\left(Y_1 + Y_2 + \dots + Y_r = m\right) \qquad m \geq r$$

horrible way is to compute it directly

$$= \sum_{\substack{i_1 + i_2 + \dots + i_r = m \\ i_\ell \geq 1}} P\left[Y_1 = i_1, Y_2 = i_2, \dots, Y_r = i_r\right]$$

$$P(Y_1 + Y_2 = m) = \sum_{j} P(Y_1 = j, Y_2 = m - j)$$

Induction Technique:

$Y_1 + Y_2 \sim NB(2, p)$

Assume: $Y_1 + \cdots + Y_{r-1} \sim NB(r-1, p)$

$(\underbrace{Y_1 + \cdots + Y_{r-1}}_{S_{r-1}}) + \underline{Y_r}$

Reproductive Property of the Gamma distribution:

My favored parametrization:

$X \sim \Gamma(\alpha, \lambda)$ if

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \mathbb{1}\{x > 0\}$$

where $\alpha > 0, \lambda > 0$.

CB:

$X \sim \Gamma(\alpha, \beta)$ if:

$$f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}} x^{\alpha-1} \mathbb{1}\{x > 0\}$$

Reproductive property:

If $X_1, X_2, \ldots, X_n$ are independent

and $X_i \sim \Gamma(\alpha_i, \beta)$,

then $S_n = \sum_{i=1}^{n} X_i \sim \Gamma\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$

Can show directly using m.g.f.'s.