# CHAPTER 1: DISTRIBUTION THEORY

## Basic Concepts

- Random Variables (R.V.)
  - discrete random variable: probability mass function
  - continuous random variable: probability density function

- *"Formal but Scary"* Definition of Random Variables
  - R.V. are some *measurable functions* from a *probability measure space* to real space;
  - probability is some non-negative value assigned to sets of a $\sigma$-*field*;
  - probability mass $\equiv$ the *Radon-Nykodym derivative* of the *random variable-induced measure* w.r.t. to a *counting measure*;
  - probability density function $\equiv$ the derivative w.r.t. *Lebesgue measure.*

- Descriptive quantities of univariate distribution
  - cumulative distribution function: $P(X \leq x)$
  - moments (mean): $E[X^k]$
  - quantiles
  - mode
  - centralized moments (variance): $E[(X-\mu)^k]$
  - the skewness: $E[(X-\mu)^3]/Var(X)^{3/2}$
  - the kurtosis: $E[(X-\mu)^4]/Var(X)^2$

- Characteristic function (c.f.)
  - $\phi_X(t) = E[\exp\{itX\}]$
  - c.f. uniquely determines the distribution

$$F_X(b) - F_X(a) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt$$

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt.$$

- Descriptive quantities of multivariate R.V.s
  - $Cov(X, Y), \quad corr(X, Y)$
  - $f_{X|Y}(x|y) = f_{X,Y}(x, y)/f_Y(y)$
  - $E[X|Y] = \int x f_{X|Y}(x|y) dx$
  - Independence:
    $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$,
    $f_{X,Y}(x, y) = f_X(x) f_Y(y)$

- Equalities of double expectations

$$E[X] = E[E[X|Y]]$$
$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

- Useful when the distribution of $X$ given $Y$ is simple!

# Examples of Discrete Distributions

- Binomial distribution
  - $Binomial(n, p)$:
    $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, ..., n$
  - $E[S_n] = np, Var(S_n) = np(1-p)$

$$\phi_X(t) = (1 - p + pe^{it})^n$$

  - $Binomial(n_1, p) + Binomial(n_2, p)$

$$\sim Binomial(n_1 + n_2, p)$$

- Negative Binomial distribution
  - $P(W_m = k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}$

    $$k = m, m+1, ...$$

  - $E[W_m] = m/p, Var(W_m) = m/p^2 - m/p$
  - Neg-Binomial$(m_1, p)$ + Neg-Binomial$(m_2, p)$

    $$\sim \text{ Neg-Binomial}(m_1 + m_2, p)$$

- Hypergeometric distribution
  - $P(S_n = k) = \binom{M}{k}\binom{N-M}{n-k}/\binom{N}{n}, \quad k = 0, 1, .., n$
  - $E[S_n] = Mn/(M + N)$
  - $Var(S_n) = \frac{nMN(M+N-n)}{(M+N)^2(M+N-1)}$

- Poisson distribution
  - $P(X = k) = \lambda^k e^{-\lambda}/k!, \quad k = 0, 1, 2, ...$
  - $E[X] = Var(X) = \lambda,$

$$\phi_X(t) = \exp\{-\lambda(1 - e^{it})\}$$

  - $Poisson(\lambda_1) + Poisson(\lambda_2) \sim Poisson(\lambda_1 + \lambda_2)$

- Poisson vs Binomial
  - $X_1 \sim Poisson(\lambda_1), \quad X_2 \sim Poisson(\lambda_2)$
  - $X_1 | X_1 + X_2 = n \sim Binomial(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$
  - if $X_{n1}, ..., X_{nn}$ are i.i.d Bernoulli($p_n$) and $np_n \to \lambda$, then

$$P(S_n = k) = \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

- Multinomial Distribution
  - $N_l, 1 \leq l \leq k$ counts the number of times that $\{Y_1, ..., Y_n\}$ fall into $B_l$
  - $P(N_1 = n_1, ..., N_k = n_k) = \binom{n}{n_1, ..., n_k} p_1^{n_1} \cdots p_k^{n_k}$

$$n_1 + ... + n_k = n$$

  - the covariance matrix for $(N_1, ..., N_k)$

$$n \begin{pmatrix} p_1(1-p_1) & \ldots & -p_1 p_k \\ \vdots & \ddots & \vdots \\ -p_1 p_k & \ldots & p_k(1-p_k) \end{pmatrix}$$

# Examples of Continuous Distribution

- Uniform distribution
  - $Uniform(a, b)$: $f_X(x) = I_{[a,b]}(x)/(b-a)$
  - $E[X] = (a+b)/2$ and $Var(X) = (b-a)^2/12$

- Normal distribution
  - $N(\mu, \sigma^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$
  - $E[X] = \mu$ and $Var(X) = \sigma^2$
  - $\phi_X(t) = \exp\{it\mu - \sigma^2 t^2/2\}$

- Gamma distribution
  - $Gamma(\theta, \beta)$: $f_X(x) = \frac{1}{\beta^\theta \Gamma(\theta)} x^{\theta-1} \exp\{-\frac{x}{\beta}\}, \quad x > 0$
  - $E[X] = \theta\beta$ and $Var(X) = \theta\beta^2$
  - $\theta = 1$ equivalent to $Exp(\beta)$
  - $\theta = n/2, \beta = 2$ equivalent to $\chi_n^2$

- Cauchy distribution
  - $Cauchy(a,b)$: $f_X(x) = \frac{1}{b\pi\{1+(x-a)^2/b^2\}}$
  - $E[|X|] = \infty$, $\phi_X(t) = \exp\{iat - |bt|\}$
  - often used as counter example

# Algebra of Random Variables

- Assumption
  - $X$ and $Y$ are independent and $Y > 0$
  - we are interested in d.f. of $X + Y, XY, X/Y$

- Summation of $X$ and $Y$
  - Derivation:

$$\begin{aligned} F_{X+Y}(z) &= E[I(X+Y \le z)] = E_Y[E_X[I(X \le z - Y)|Y]] \\ &= E_Y[F_X(z-Y)] = \int F_X(z-y)dF_Y(y) \end{aligned}$$

  - Convolution formula:

$$F_{X+Y}(z) = \int F_Y(z-x)dF_X(x) \equiv F_X * F_Y(z)$$

$$f_X * f_Y(z) \equiv \int f_X(z-y)f_Y(y)dy = \int f_Y(z-x)f_X(x)dx$$

- Product and quotient of $X$ and $Y$ $(Y > 0)$
  - $F_{XY}(z) = E[E[I(XY \leq z)|Y]] = \int F_X(z/y)dF_Y(y)$

$$f_{XY}(z) = \int f_X(z/y)/y f_Y(y)dy$$

  - $F_{X/Y}(z) = E[E[I(X/Y \leq z)|Y]] = \int F_X(yz)dF_Y(y)$

$$f_{X/Y}(z) = \int f_X(yz)y f_Y(y)dy$$

- Application of formulae
  - $N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
  - $Gamma(r_1, \theta) + Gamma(r_2, \theta) \sim Gamma(r_1 + r_2, \theta)$
  - $Poisson(\lambda_1) + Poisson(\lambda_2) \sim Poisson(\lambda_1 + \lambda_2)$
  - Negative Binomial$(m_1, p)$ + Negative Binomial$(m_2, p)$

    $\sim$ Negative Binomial$(m_1 + m_2, p)$

- Summation of R.V.s using c.f.
  - Result: if $X$ and $Y$ are independent, then
    $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$
  - Example: $X$ and $Y$ are normal with
    $\phi_X(t) = \exp\{i\mu_1 t - \sigma_1^2 t^2/2\}$,
    $\phi_Y(t) = \exp\{i\mu_2 t - \sigma_2^2 t^2/2\}$

$$\Rightarrow \phi_{X+Y}(t) = \exp\{i(\mu_1 + \mu_2)t - (\sigma_1^2 + \sigma_2^2)t^2/2\}$$

- Further examples of special distribution
  - assume $X \sim N(0,1)$, $Y \sim \chi^2_m$ and $Z \sim \chi^2_n$ are independent;
  - $$\frac{X}{\sqrt{Y/m}} \sim \text{ Student's } \ t(m),$$
    $$\frac{Y/m}{Z/n} \sim \text{Snedecor's } \ F_{m,n},$$
    $$\frac{Y}{Y+Z} \sim \text{ Beta}(m/2, n/2).$$

- Densities of $t-$, $F-$ and $Beta-$ distributions
  - $f_{t(m)}(x) = \frac{\Gamma((m+1)/2)}{\sqrt{\pi m}\Gamma(m/2)} \frac{1}{(1+x^2/m)^{(m+1)/2}} I_{(-\infty,\infty)}(x)$
  - $f_{F_{m,n}}(x) = \frac{\Gamma(m+n)/2}{\Gamma(m/2)\Gamma(n/2)} \frac{(m/n)^{m/2} x^{m/2-1}}{(1+mx/n)^{(m+n)/2}} I_{(0,\infty)}(x)$
  - $f_{Beta(a,b)}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} I(x \in (0,1))$

- Exponential vs Beta- distributions
  - assume $Y_1, ..., Y_{n+1}$ are i.i.d Exp($\theta$);
  - $Z_i = \frac{Y_1+...+Y_i}{Y_1+...+Y_{n+1}} \sim \text{Beta}(i, n-i+1)$;
  - $(Z_1, \ldots, Z_n)$ has the same joint distribution as that of the order statistics $(\xi_{n:1}, ..., \xi_{n:n})$ of $n$ Uniform(0,1) r.v.s.

# Transformation of Random Vector

**Theorem 1.3** Suppose that $X$ is $k$-dimension random vector with density function $f_X(x_1, ..., x_k)$. Let $g$ be a one-to-one and continuously differentiable map from $R^k$ to $R^k$. Then $Y = g(X)$ is a random vector with density function

$$f_X(g^{-1}(y_1, ..., y_k))|J_{g^{-1}}(y_1, ..., y_k)|,$$

where $g^{-1}$ is the inverse of $g$ and $J_{g^{-1}}$ is the Jacobian of $g^{-1}$.

- Example
  - let $R^2 \sim Exp\{2\}, R > 0$ and $\Theta \sim Uniform(0, 2\pi)$ be independent;
  - $X = R\cos\Theta$ and $Y = R\sin\Theta$ are two independent standard normal random variables;
  - it can be applied to simulate normally distributed data.

# Multivariate Normal Distribution

- Definition

$Y = (Y_1, ..., Y_n)'$ is said to have a multivariate normal distribution with mean vector $\mu = (\mu_1, ..., \mu_n)'$ and non-degenerate covariance matrix $\Sigma_{n\times n}$ if

$$f_Y(y_1, ..., y_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\}$$

- Characteristic function

$$
\begin{aligned}
& \phi_Y(t) \\
= \quad & E[e^{it'Y}] = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \int \exp\{it'y - \frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\} dy \\
= \quad & (\sqrt{2\pi})^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \int \exp\{-\frac{y'\Sigma^{-1}y}{2} + (it + \Sigma^{-1}\mu)'y - \frac{\mu'\Sigma^{-1}\mu}{2}\} dy \\
= \quad & \frac{\exp\{-\mu'\Sigma^{-1}\mu/2\}}{(\sqrt{2\pi})^{n/2}|\Sigma|^{1/2}} \int \exp\left\{-\frac{1}{2}(y - \Sigma it - \mu)'\Sigma^{-1}(y - \Sigma it - \mu)\right. \\
& \qquad \left. + (\Sigma it + \mu)'\Sigma^{-1}(\Sigma it + \mu)/2\right\} dy \\
= \quad & \exp\left\{-\mu'\Sigma^{-1}\mu/2 + \frac{1}{2}(\Sigma it + \mu)'\Sigma^{-1}(\Sigma it + \mu)\right\} \\
= \quad & \exp\{it'\mu - \frac{1}{2}t'\Sigma t\}.
\end{aligned}
$$

- Conclusions
  - multivariate normal distribution is uniquely determined by $\mu$ and $\Sigma$
  - for standard multivariate normal distribution, $\phi_X(t) = \exp\{-t't/2\}$
  - the moment generating function for $Y$ is $\exp\{t'\mu + t'\Sigma t/2\}$

- Linear transformation of normal r.v.

**Theorem 1.4** If $Y = A_{n\times k}X_{k\times 1}$ where $X \sim N(0, I)$ (standard multivariate normal distribution), then $Y$'s characteristic function is given by

$$\phi_Y(t) = \exp\{-t'\Sigma t/2\}, \quad t = (t_1, ..., t_n) \in R^k$$

and $rank(\Sigma) = rank(A)$. Conversely, if $\phi_Y(t) = \exp\{-t'\Sigma t/2\}$ with $\Sigma_{n\times n} \geq 0$ of rank $k$, then

$$Y = A_{n\times k}X_{k\times 1} \text{ with } rank(A) = k \text{ and } X \sim N(0, I).$$

# Proof

$Y = AX$ and $X \sim N(0, I) \Rightarrow$

$$\phi_Y(t) = E[\exp\{it'(AX)\}] = E[\exp\{i(A't)'X\}]$$

$$= \exp\{-(A't)'(A't)/2\} = \exp\{-t'AA't/2\}$$

$\Rightarrow Y \sim N(0, AA')$.

Note that $rank(AA') = rank(A)$.

Conversely, suppose $\phi_Y(t) = \exp\{-t'\Sigma t/2\}$. There exists $O$, an orthogonal matrix,

$$\Sigma = O'DO, \quad D = diag((d_1, ..., d_k, 0, ..., 0)'), \quad d_1, ..., d_k > 0.$$

Define $Z = OY$

$$\Rightarrow \phi_Z(t) = E[\exp\{it'(OY)\}] = E[\exp\{i(O't)'Y\}]$$

$$= \exp\{-\frac{(O't)'\Sigma(O't)}{2}\} = \exp\{-d_1t_1^2/2 - ... - d_kt_k^2/2\}.$$

$\Rightarrow$ $Z_1, ..., Z_k$ are independent $N(0, d_1), ..., N(0, d_k)$ and $Z_{k+1}, ..., Z_n$ are zeros.

$\Rightarrow$ Let $X_i = Z_i/\sqrt{d_i}, i = 1, ..., k$ and $O' = (B_{n\times k}, C_{n\times(n-k)})$.

$$Y = O'Z = B\text{diag}\{(\sqrt{d_1}, ..., \sqrt{d_k})\}X \equiv AX.$$

Clearly, $rank(A) = k$.

- Conditional normal distributions

**Theorem 1.5** Suppose that $Y = (Y_1, ..., Y_k, Y_{k+1}, ..., Y_n)'$ has a multivariate normal distribution with mean $\mu = (\mu^{(1)'}, \mu^{(2)'})'$ and a non-degenerate covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then
(i) $(Y_1, ..., Y_k)' \sim N_k(\mu^{(1)}, \Sigma_{11})$.
(ii) $(Y_1, ..., Y_k)'$ and $(Y_{k+1}, ..., Y_n)'$ are independent if and only if $\Sigma_{12} = \Sigma_{21} = 0$.
(iii) For any matrix $A_{m\times n}$, $AY$ has a multivariate normal distribution with mean $A\mu$ and covariance $A\Sigma A'$.

(iv) The conditional distribution of $Y^{(1)} = (Y_1, ..., Y_k)'$ given $Y^{(2)} = (Y_{k+1}, ..., Y_n)'$ is a multivariate normal distribution given as

$$Y^{(1)}|Y^{(2)} \sim N_k(\mu^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)}-\mu^{(2)}), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

# Proof

Prove (iii) first. $\phi_Y(t) = \exp\{it'\mu - t'\Sigma t/2\}$.

$$\Rightarrow \qquad \phi_{AY}(t) = E[e^{it'AY}] = E[e^{i(A't)'Y}]$$

$$= \exp\{i(A't)'\mu - (A't)'\Sigma(A't)/2\}$$

$$= \exp\{it'(A\mu) - t'(A\Sigma A')t/2\}$$

$$\Rightarrow \qquad AY \sim N(A\mu, A\Sigma A').$$

Prove (i).

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} = \left( I_{k\times k} \quad 0_{k\times(n-k)} \right) Y.$$

$\Rightarrow$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \sim N\left( \left( I_{k\times k} \quad 0_{k\times(n-k)} \right) \mu, \right.$$

$$\left. \left( I_{k\times k} \quad 0_{k\times(n-k)} \right) \Sigma \left( I_{k\times k} \quad 0_{k\times(n-k)} \right)' \right).$$

Prove (ii). $t = (t^{(1)}, t^{(2)})'$

$$\phi_Y(t) = \exp\left[ it^{(1)'}\mu^{(1)} + it^{(2)'}\mu^{(2)} - \frac{1}{2}\left\{ t^{(1)'}\Sigma_{11}t^{(1)} + 2t^{(1)'}\Sigma_{12}t^{(2)} + t^{(2)'}\Sigma_{22}t^{(2)} \right\} \right].$$

$\Rightarrow$ $t^{(1)}$ and $t^{(2)}$ are separable iff $\Sigma_{12} = 0$.

(ii) implies that two normal random variables are independent if and only if their covariance is zero.

Prove (iv). Consider

$$Z^{(1)} = Y^{(1)} - \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)} - \mu^{(2)}).$$

$\Rightarrow$ $Z^{(1)} \sim N(0, \Sigma_Z)$

$$\Sigma_Z = Cov(Y^{(1)}, Y^{(1)}) - 2\Sigma_{12}\Sigma_{22}^{-1}Cov(Y^{(2)}, Y^{(1)})$$
$$+\Sigma_{12}\Sigma_{22}^{-1}Cov(Y^{(2)}, Y^{(2)})\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

$$Cov(Z^{(1)}, Y^{(2)}) = Cov(Y^{(1)}, Y^{(2)}) - \Sigma_{12}\Sigma_{22}^{-1}Cov(Y^{(2)}, Y^{(2)}) = 0.$$

$\Rightarrow$ $Z^{(1)}$ is independent of $Y^{(2)}$.

$\Rightarrow$ The conditional distribution $Z^{(1)}$ given $Y^{(2)}$ is the same as the unconditional distribution of $Z^{(1)}$

$\Rightarrow$ $Z^{(1)}|Y^{(2)} \sim Z^{(1)} \sim N(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

- Remark on Theorem 1.5
  - The constructed $Z$ in (iv) has a geometric interpretation using the projection of $(Y^{(1)} - \mu^{(1)})$ on the normalized variable $\Sigma_{22}^{-1/2}(Y^{(2)} - \mu^{(2)})$.

- Example
  - $X$ and $U$ are independent, $X \sim N(0, \sigma_x^2)$ and $U \sim N(0, \sigma_u^2)$
  - $Y = X + U$ (measurement error)
  - the covariance of $(X, Y)$: $\begin{pmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}$
  - $X|Y \sim N(\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2} Y, \sigma_x^2 - \frac{\sigma_x^4}{\sigma_x^2+\sigma_u^2}) \equiv N(\lambda Y, \sigma_x^2(1-\lambda))$

$$\lambda = \sigma_x^2/\sigma_y^2 \text{ (reliability coefficient)}$$

- Quadratic form of normal random variables
  - $N(0,1)^2 + ... + N(0,1)^2 \sim \chi_n^2 \equiv Gamma(n/2, 2)$.
  - If $Y \sim N_n(0, \Sigma)$ with $\Sigma > 0$, then $Y'\Sigma^{-1}Y \sim \chi_n^2$.
  - Proof:

    $Y = AX$ where $X \sim N(0, I)$ and $\Sigma = AA'$.

$$Y'\Sigma^{-1}Y = X'A'(AA')^{-1}AX = X'X \sim \chi_n^2.$$

- Noncentral Chi-square distribution
  - assume $X \sim N(\mu, I)$
  - $Y = X'X \sim \chi^2_n(\delta)$ and $\delta = \mu'\mu$.
  - $Y$'s density:

$$f_Y(y) = \sum_{k=0}^{\infty} \frac{\exp\{-\delta/2\}(\delta/2)^k}{k!} g(y; (2k+n)/2, 1/2).$$

- Additional notes on normal distributions
  - noncentral $t$-distribution: $N(\delta, 1)/\sqrt{\chi_n^2/n}$
    noncentral $F$-distribution: $\chi_n^2(\delta)/n/(\chi_m^2/m)$
  - how to calculate $E[X'AX]$ where $X \sim N(\mu, \Sigma)$:

$$\begin{aligned} E[X'AX] &= E[tr(X'AX)] = E[tr(AXX')] \\ &= tr(AE[XX']) = tr(A(\mu\mu' + \Sigma)) \\ &= \mu' A\mu + tr(A\Sigma) \end{aligned}$$

# Families of Distributions

- *location-scale* family
  - $aX + b$: $f_X((x-b)/a)/a, \quad a > 0, b \in R$
  - mean $aE[X] + b$ and variance $a^2 Var(X)$
  - examples: normal distribution, uniform distribution, gamma distributions etc.

- Exponential family
  - examples: binomial, poisson distributions for discrete variables and normal distribution, gamma distribution, Beta distribution for continuous variables
  - has a general expression of densities
  - possesses nice statistical properties

- Form of densities
  - $\{P_\theta\}$, is said to form an $s$-parameter exponential family:

$$p_\theta(x) = \exp\left\{\sum_{k=1}^{s} \eta_k(\theta)T_k(x) - B(\theta)\right\} h(x)$$

$$\exp\{B(\theta)\} = \int \exp\{\sum_{k=1}^{s} \eta_k(\theta)T_k(x)\}h(x)d\mu(x) < \infty$$

  - if $\{\eta_k(\theta)\} = \theta$, the above form is called the canonical form of the exponential family.

- Examples
  - $X_1, ..., X_n \sim N(\mu, \sigma^2)$:

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

  - $X \sim Binomial(n, p)$:

$$\exp\{x \log\frac{p}{1-p} + n\log(1-p)\}\binom{n}{x}$$

  - $X \sim Poisson(\lambda)$:

$$P(X = x) = \exp\{x \log\lambda - \lambda\}/x!$$

- Moment generating function (MGF)
  - MGF for $(T_1, ..., T_s)$:

$$M_T(t_1, ..., t_s) = E\left[\exp\{t_1 T_1 + ... + t_s T_s\}\right]$$

  - the coefficients in the Taylor expansion of $M_T$ correspond to the moments of $(T_1, ..., T_s)$

- Calculate MGF in the exponential family

**Theorem 1.6** Suppose the densities of an exponential family can be written as the canonical form

$$\exp\{\sum_{k=1}^{s} \eta_k T_k(x) - A(\eta)\} h(x),$$

where $\eta = (\eta_1, ..., \eta_s)'$. Then for $t = (t_1, ..., t_s)'$,

$$M_T(t) = \exp\{A(\eta + t) - A(\eta)\}.$$

# Proof

$\exp\{A(\eta)\} = \int \exp\{\sum_{k=1}^{s} \eta_i T_i(x)\} h(x) d\mu(x)$.

$\Rightarrow$

$$M_T(t) = E\left[\exp\{t_1 T_1 + ... + t_s T_s\}\right]$$

$$= \int \exp\{\sum_{k=1}^{s} (\eta_i + t_i) T_i(x) - A(\eta)\} h(x) d\mu(x)$$

$$= \exp\{A(\eta + t) - A(\eta)\}.$$

- Cumulant generating function (CGF)
  - $K_T(t_1, ..., t_s) = \log M_T(t_1, ..., t_s) = A(\eta + t) - A(\eta)$
  - the coefficients in the Taylor expansion are called the cumulants for $(T_1, ..., T_s)$
  - the first two cumulants are the mean and variance

- Examples revisited
  - Normal distribution: $\eta = \mu/\sigma^2$ and

$$A(\eta) = \frac{1}{2\sigma^2}\mu^2 = \eta^2\sigma^2/2$$

$\Rightarrow$

$$M_T(t) = \exp\{\frac{\sigma^2}{2}((\eta + t)^2 - \eta^2)\} = \exp\{\mu t + t^2\sigma^2/2\}.$$

From the Taylor expansion, for $X \sim N(0, \sigma^2)$,

$E[X^{2r+1}] = 0, E[X^{2r}] = 1 \cdot 2 \cdots (2r-1)\sigma^{2r}, r = 1, 2, ...$

– gamma distribution has a canonical form

$$\exp\{-x/b + (a-1)\log x - \log(\Gamma(a)b^a)\}I(x > 0).$$

$\Rightarrow$ $$\eta = -1/b, T = X$$

$$A(\eta) = \log(\Gamma(a)b^a) = a\log(-1/\eta) + \log\Gamma(a).$$

$\Rightarrow$

$$M_X(t) = \exp\{a\log\frac{\eta}{\eta+t}\} = (1-bt)^{-a}.$$

$$E[X] = ab, E[X^2] = ab^2 + (ab)^2, ...$$

*READING MATERIALS*: Lehmann and Casella, Sections 1.4 and 1.5