Lecture 7: Moments, inequalities, m.g.f. and ch.f.

If EX^k is finite, where k is a positive integer, EX^k is called the kth *moment* of X or P_X.
If $E|X|^a < \infty$ for some real number a, $E|X|^a$ is called the ath *absolute moment* of X or P_X.
If $\mu = EX$ and $E(X-\mu)^k$ are finite for a positive integer k, $E(X-\mu)^k$ is called the kth *central moment* of X or P_X.
Variance: $E(X-EX)^2$

$X = (X_1, ..., X_k)$, $EX = (EX_1, ..., EX_k)$
$M = (M_{ij})$, $EM = (EM_{ij})$
Covariance matrix: $\text{Var}(X) = E(X-EX)(X-EX)^\tau$
The (i,j)th element of $\text{Var}(X)$, $i \neq j$, is $E(X_i - EX_i)(X_j - EX_j)$, which is called the *covariance* of X_i and X_j and is denoted by $\text{Cov}(X_i, X_j)$.
$\text{Var}(X)$ is nonnegative definite
$[\text{Cov}(X_i, X_j)]^2 \leq \text{Var}(X_i)\text{Var}(X_j), \quad i \neq j$
If $\text{Cov}(X_i, X_j) = 0$, then X_i and X_j are uncorrelated
Independence implies uncorrelation, not converse
If $Y = c^\tau X$, $c \in \mathcal{R}^k$, and X is a random k-vector, $EY = c^\tau EX$ and $\text{Var}(Y) = c^\tau \text{Var}(X)c$.

Three useful inequalities
Cauchy-Schwartz inequality: $[E(XY)]^2 \leq EX^2 EY^2$ for random variables X and Y
Jensen's inequality: $f(EX) \leq Ef(X)$ for a random vector X and convex function f ($f'' \geq 0$)
Chebyshev's inequality: Let X be a random variable and φ a nonnegative and nondecreasing function on $[0,\infty)$ satisfying $\varphi(-t) = \varphi(t)$. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X)dP \leq E\varphi(X)$$

Example 1.18. If X is a nonconstant positive random variable with finite mean, then

$$(EX)^{-1} < E(X^{-1}) \quad \text{and} \quad E(\log X) < \log(EX),$$

since t^{-1} and $-\log t$ are convex functions on $(0,\infty)$. Let f and g be positive integrable functions on a measure space with a σ-finite measure ν. If $\int f d\nu \geq \int g d\nu > 0$, we want to show that

$$\int f \log\left(\frac{f}{g}\right) d\nu \geq 0.$$

Let $h = f / \int f d\nu$. Then h is a p.d.f. w.r.t. ν. Let $Y = g/f$ be a random variable defined on the probability space with P being the probability with p.d.f. h. By Jensen's inequality, $E\log(g/f) \leq \log(E(g/f))$. Note that

$$\log(E(g/f)) = \log\left(\int \frac{g}{f} h d\nu\right) = \log\left(\frac{\int g d\nu}{\int f d\nu}\right) \leq 0$$

and

$$E\log(g/f) = \int \log\left(\frac{g}{f}\right) h d\nu = \int \log\left(\frac{g}{f}\right) f d\nu \Big/ \int f d\nu$$

Moment generating and characteristic functions

Definition 1.5. Let X be a random k-vector.

(i) The *moment generating function* (m.g.f.) of X or P_X is defined as

$$\psi_X(t) = Ee^{t^\tau X}, \quad t \in \mathcal{R}^k.$$

(ii) The *characteristic function* (ch.f.) of X or P_X is defined as

$$\phi_X(t) = Ee^{\sqrt{-1}t^\tau X} = E[\cos(t^\tau X)] + \sqrt{-1}\, E[\sin(t^\tau X)], \quad t \in \mathcal{R}^k$$

If the m.g.f. is finite in a neighborhood of $0 \in \mathcal{R}^k$, then $\phi_X(t)$ can be obtained by replacing t in $\psi_X(t)$ by $\sqrt{-1}t$

If $Y = A^\tau X + c$, where A is a $k \times m$ matrix and $c \in \mathcal{R}^m$, it follows from Definition 1.5 that

$$\psi_Y(u) = e^{c^\tau u}\psi_X(Au) \quad \text{and} \quad \phi_Y(u) = e^{\sqrt{-1}c^\tau u}\phi_X(Au), \quad u \in \mathcal{R}^m$$

$X = (X_1, ..., X_k)$ with m.g.f. ψ_X finite in a neighborhood of 0

$$\psi_X(t) = \sum_{(r_1,\ldots,r_k)} \frac{\mu_{r_1,\ldots,r_k} t_1^{r_1} \cdots t_k^{r_k}}{r_1! \cdots r_k!} \qquad \mu_{r_1,\ldots,r_k} = E(X_1^{r_1} \cdots X_k^{r_k})$$

Special case of $k = 1$:

$$\psi_X(t) = \sum_{i=0}^{\infty} \frac{E(X^i)t^i}{i!}$$

Consequently,

$$E(X_1^{r_1} \cdots X_k^{r_k}) = \left.\frac{\partial^{r_1+\cdots+r_k}\psi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_k}}\right|_{t=0} \qquad E(X^i) = \psi^{(i)}(0) = \left.\frac{d\psi_X^i(t)}{dt^i}\right|_{t=0}$$

$$\left.\frac{\partial \psi_X(t)}{\partial t}\right|_{t=0} = EX, \qquad \left.\frac{\partial^2 \psi_X(t)}{\partial t \partial t^\tau}\right|_{t=0} = E(XX^\tau)$$

If $0 < \psi_X(t) < \infty$, then $\kappa_X(t) = \log \psi_X(t)$ is called the *cumulant generating function* of X or P_X.

If ψ_X is not finite and $E|X_1^{r_1} \cdots X_k^{r_k}| < \infty$ for some nonnegative integers $r_1, ..., r_k$, then

$$\left.\frac{\partial^{r_1+\cdots+r_k}\phi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_k}}\right|_{t=0} = (-1)^{(r_1+\cdots+r_k)/2} E(X_1^{r_1} \cdots X_k^{r_k})$$

$$\left.\frac{\partial \phi_X(t)}{\partial t}\right|_{t=0} = \sqrt{-1}EX, \qquad \left.\frac{\partial^2 \phi_X(t)}{\partial t \partial t^\tau}\right|_{t=0} = -E(XX^\tau), \qquad \phi_X^{(i)}(0) = (-1)^{i/2}E(X^i)$$

Example: a random variable X has finite $E(X^k)$ for $k = 1, 2...$ but $\psi_X(t) = \infty$, $t \neq 0$
P_n: the probability measure for $N(0, n)$ with p.d.f. f_n, $n = 1, 2, ...$
$P = \sum_{n=1}^{\infty} 2^{-n} P_n$ is a probability measure with Lebesgue p.d.f. $\sum_{n=1}^{\infty} 2^{-n} f_n$ (Exercise 35)
Let X be a random variable having distribution P.
It follows from Fubini's theorem that X has finite moments of any order; for even k,

$$E(X^k) = \int x^k dP = \int \sum_{n=1}^{\infty} x^k 2^{-n} dP_n = \sum_{n=1}^{\infty} 2^{-n} \int x^k dP_n = \sum_{n=1}^{\infty} 2^{-n}(k-1)(k-3)\cdots 1 n^{k/2} < \infty$$

and $E(X^k) = 0$ for odd k.
By Fubini's theorem,

$$\psi_X(t) = \int e^{tx} dP = \sum_{n=1}^{\infty} 2^{-n} \int e^{tx} dP_n = \sum_{n=1}^{\infty} 2^{-n} e^{nt^2/2} = \infty \quad t \neq 0$$

Since the ch.f. of $N(0, n)$ is $e^{-nt^2/2}$,

$$\phi_X(t) = \int e^{\sqrt{-1}tx} dP = \sum_{n=1}^{\infty} 2^{-n} \int e^{\sqrt{-1}tx} dP_n = \sum_{n=1}^{\infty} 2^{-n} e^{-nt^2/2} = (2e^{t^2/2} - 1)^{-1}$$

(Fubini's theorem)
Hence, the moments of X can be obtained by differentiating ϕ_X
For example, $\phi'_X(0) = 0$ and $\phi''_X(0) = -2$, which shows that $EX = 0$ and $EX^2 = 2$.

Theorem 1.6. (Uniqueness). Let X and Y be random k-vectors.
(i) If $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathcal{R}^k$, then $P_X = P_Y$.
(ii) If $\psi_X(t) = \psi_Y(t) < \infty$ for all t in a neighborhood of 0, then $P_X = P_Y$.

Another useful result: For independent X and Y,

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t) \quad \text{and} \quad \phi_{X+Y}(t) = \phi_X(t)\psi_Y(t), \quad t \in \mathcal{R}^k$$

Example 1.20. Let X_i, $i = 1, ..., k$, be independent random variables and X_i have the gamma distribution $\Gamma(\alpha_i, \gamma)$ (Table 1.2), $i = 1, ..., k$. From Table 1.2, X_i has the m.g.f. $\psi_{X_i}(t) = (1 - \gamma t)^{-\alpha_i}$, $t < \gamma^{-1}$, $i = 1, ..., k$. Then, the m.g.f. of $Y = X_1 + \cdots + X_k$ is equal to $\psi_Y(t) = (1 - \gamma t)^{-(\alpha_1 + \cdots + \alpha_k)}$, $t < \gamma^{-1}$. From Table 1.2, the gamma distribution $\Gamma(\alpha_1 + \cdots + \alpha_k, \gamma)$ has the m.g.f. $\psi_Y(t)$ and, hence, is the distribution of Y (by Theorem 1.6).

A random vector X is symmetric about 0 iff X and $-X$ have the same distribution
Show that: X is symmetric about 0 if and only if its ch.f. ϕ_X is real-valued.
If X and $-X$ have the same distribution, then by Theorem 1.6, $\phi_X(t) = \phi_{-X}(t)$.
But $\phi_{-X}(t) = \phi_X(-t)$. Then $\phi_X(t) = \phi_X(-t)$.
Note that $\sin(-t^\tau X) = -\sin(t^\tau X)$ and $\cos(t^\tau X) = \cos(-t^\tau X)$
Hence $E[\sin(t^\tau X)] = 0$ and, thus, ϕ_X is real-valued.
Conversely, if ϕ_X is real-valued, then $\phi_X(t) = E[\cos(t^\tau X)]$ and $\phi_{-X}(t) = \phi_X(-t) = \phi_X(t)$.
By Theorem 1.6, X and $-X$ must have the same distribution.