2.3 Moments, mean and variance

Definition. Let X be a r.v.

(i) The k-th **moment** of X is $\mu'_k = E(X^k)$ if $E(|X|^k) < \infty$.

(ii) The k-th **central moment** of X is $\mu_k = E\{(X - EX)^k\}$ if $E(|X|^k) < \infty$.

(iii) Important special cases:
mean: $\mu_X = \mu'_1 = E(X)$,
variance: $var(X) = \mu_2 = E\{(X - \mu_X)^2\} = \sigma_X^2$,
standard deviation: $\sigma_X = \sqrt{var(X)}$.

Theorem.
(i) $\sigma_X^2 = E(X^2) - \{E(X)\}^2 = E\{X(X-1)\} - E(X)\{E(X) - 1\}$,
(ii) $Y = aX + b \Rightarrow \mu_Y = a\mu_X + b$,
(iii) $Y = aX + b \Rightarrow \sigma_Y^2 = a^2\sigma_X^2$.

Proof. (i) $\sigma_X^2 = E\{(X - \mu_X)^2\} = E(X^2 + \mu_X^2 - 2X\mu_X) = E(X^2) + \mu_X^2 - 2\mu_X^2 = E(X^2) - \{E(X)\}^2$. Second statement: exercise.
(ii) Linearity.
(iii) $\sigma_Y^2 = E[\{aX + b - E(aX + b)\}^2] = E\{(aX - a\mu_X)^2\} = a^2 E\{(X - \mu_X)^2\} = a^2\sigma_X^2$. □

Example. Find $var(Y)$ for $Y \sim$ binomial(n, p). We know from a previous example: $\mu_Y = np$, $E\{Y(Y-1)\} = n(n-1)p^2$. The above theorem thus gives $var(Y) = E\{Y(Y-1)\} - E(Y)\{E(Y) - 1\} = n(n-1)p^2 - np(np-1) = n^2p^2 - np^2 - n^2p^2 + np = np(1-p)$.

Example: standard normal distribution. Let Z be standard normally distributed, i.e. with pdf

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

We use $\frac{d}{dz} f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}(-\frac{1}{2})2z = -z f_Z(z)$ to compute the mean and the variance of Z:
$\mu_Z = E(Z) = \int_{-\infty}^{\infty} z f_Z(z)\, dz = -f_Z(z) \mid_{-\infty}^{\infty} = 0 - 0 = 0$,
$E(Z^2) = \int_{-\infty}^{\infty} z\{z f_Z(z)\}\, dz = -z f_Z(z) \mid_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f_Z(z)\, dz = 0 + 1 = 1$,
$\sigma_Z^2 = var(Z) = E(Z^2) - \mu_Z^2 = 1 - 0 = 1$.

Now consider a normally distributed r.v. X with parameters μ and σ (or σ^2), $X = \mu + \sigma Z$. Then $\mu_X = \mu + \sigma\mu_Z = \mu$ and $\sigma_X^2 = \sigma^2\sigma_Z^2 = \sigma^2$, i.e. the parameters of the $N(\mu, \sigma^2)$ distribution are its mean and variance.

Theorem.
(i) $t > 0, m > 0 \Rightarrow P(|X| > t) \leq \dfrac{E(|X|^m)}{t^m}$ (**Markov** inequality);

(ii) $t > 0 \Rightarrow P(|X - \mu_X| > t) \leq \dfrac{\sigma_X^2}{t^2}$ (**Chebyshev** inequality).

Proof. (ii) follows from (i) with $\tilde{X} = X - \mu_X$, $m = 2$.
(i) Use $1_{|X|>t} \leq (|X|/t)^m$ for $t > 0, m > 0$ and the facts that the expectation is monotone and linear to obtain $P(|X| > t) = E(1_{|X|>t}) \leq E((|X|/t)^m) = 1/t^m E(|X|^m)$. □

Theorem. Let X be a r.v. and $m > 0$. Then

$$E(|X|^m) < \infty \;\Rightarrow\; E(|X|^k) < \infty \quad \text{for all } k \in (0, m).$$

Proof. Use $|X|^k \le 1 + |X|^m$ which holds since $|X|^k < 1$ for $|X| < 1$ and $|X|^k \le |X|^m$ for $|X| \ge 1$. Therefore, $E(|X|^k) \le E(1+|X|^m) = 1+E(|X|^m) < \infty$, $\square$

Preliminaries. We know $0 \le Var(X) = E(X-\mu)^2 = E(X^2) - (EX)^2 \Rightarrow E(X^2) \ge (EX)^2$, i.e. for the *convex* function $g(x) = x^2$ holds

$$E\{g(X)\} \ \ge \ g\{E(X)\}.$$

Theorem (Jensen's inequality). Consider a r.v. X, a convex function g on the support of X and assume that both $E(X)$ and $E\{g(X)\}$ are finite. Then $E\{g(X)\} \ge g\{E(X)\}$.

Proof. Consider the point $(E(X), g\{E(X)\})$ on the graph of g and the tangent line t through this point, i.e. $t\{E(X)\} = g\{E(X)\}$. The line t is of the form $t(x) = a+bx$ with some constants a and b. The function g is convex, i.e. the graph of $g(x)$ lies above its tangent lines. Therefore, $g(x) \ge a + bx$ and, since the expectation is monotone and linear, $E\{g(X)\} \ge a + bE(X) = t\{E(X)\} = g\{E(X)\}$. $\square$

Examples. (i) $g(x) = |x| : E|X| \ge |EX|$; (ii) $g(x) = \frac{1}{x}$ with $x > 0$: $E(\frac{1}{X}) \ge \frac{1}{EX}$.

2.4 Generating functions

Definition. Let X be a random variable.
(i) The **moment generating function** (mgf) of X is $M_X(t) = E(e^{tX})$.
(ii) The **characteristic function** (cf) of X is
$\phi_X(t) = E(e^{itX}) = E\{\cos(tX) + i\sin(tX)\} \quad (i = \sqrt{-1})$.

Note. The characteristic function is always finite, the moment generating function not necessarily. If $M_X(t) < \infty$ for all $t \in (-\delta, \delta)$ then $\phi_X(t) = M_X(it)$.

Example. Let $X \sim \text{expo}(\beta)$. The mgf is defined on $(-\infty, 1/\beta)$,

$$M_X(t) = \int_0^\infty e^{tx}\frac{1}{\beta}e^{-\frac{x}{\beta}}\,dx = \frac{1}{\beta}\frac{1}{\frac{1}{\beta}-t}\int_0^\infty (\frac{1}{\beta}-t)e^{-x(\frac{1}{\beta}-t)}\,dx = \frac{1}{\beta}\frac{1}{\frac{1}{\beta}-t} = \frac{1}{1-\beta t}.$$

Example. For $Y \sim \text{binomial}(n,p)$ the moment generating function is $M_Y(t) = \sum_{k=0}^n e^{tk}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k}(e^tp)^k(1-p)^{n-k} = (pe^t+1-p)^n$. Here we have used the **binomial formula:**

$$\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n.$$

Theorem. Consider X with mgf M_X and let $Y = aX + b$. Then Y has the mgf $M_Y(t) = e^{bt}M_X(at)$.

Proof. $E(e^{tY}) = E(e^{t(aX+b)}) = E(e^{taX})e^{tb} = M_X(at)e^{tb}$. $\square$

Example. Consider $Z \sim N(0,1)$ and $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$. The mgf of Z is computed by completing the square so that we obtain an integral of a $N(t,1)$ pdf:

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^\infty e^{tz}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\,dz = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}+tz-\frac{t^2}{2}+\frac{t^2}{2}}\,dz \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{(z-t)^2}{2}+\frac{t^2}{2}}\,dz = e^{\frac{t^2}{2}}\cdot 1 = e^{\frac{t^2}{2}}. \end{aligned}$$

Now use the previous theorem to obtain $M_X(t) = e^{\mu t}M_Z(\sigma t) = e^{\mu t + \sigma^2 t^2/2}$.