Theorem 3.3 (Chebyshev's inequality, TBp. 133)

Let X be a random variable with mean μ and variance σ^2. Then for any $t > 0$,

discrete or continuous

$$0 \leq P(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

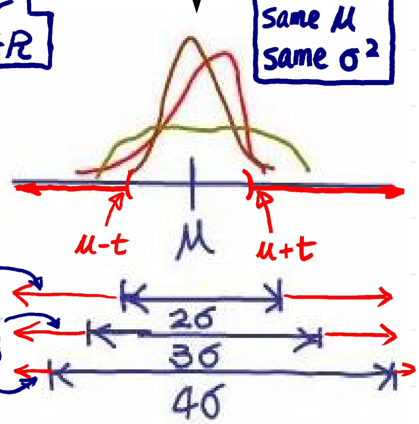
prove for continuous case, the proof for discrete case is analogous.

Proof. Let $f(x)$ be the pdf of X. Let $R = \{x : |x-\mu| \geq t\} \Leftrightarrow \frac{(x-\mu)^2}{t^2} \geq 1$. Then

$$\underset{P(R)}{P(|X-\mu| \geq t)} = \int_R 1 \cdot f(x)\,dx \leq \int_R \frac{(x-\mu)^2}{t^2} f(x)\,dx$$
$$\leq \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{t^2} f(x)\,dx = \frac{\sigma^2}{t^2}.$$

≥ 1 for $x \in R$

No other restriction on the functional form of pdf/pmf



Note.

1. Setting $t = k\sigma$ ($k = t/\sigma$) we have

"=" holds for $X = \begin{cases} -1, & \text{with prob. } 1/2k^2 \\ 0, & \text{with prob. } 1-1/k^2 \\ 1, & \text{with prob. } 1/2k^2 \end{cases}$ $\mu=0, \sigma^2 = 1/k^2$ (Ec)

This explains why σ is called standard deviation (標準差)

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

≤ 1 $\quad \leq 1/1.5^2 \quad \leq 1/4$

2. (TBp. 134) $Var(X) = 0 \Rightarrow P(X = \mu_X) = 1.$

$P(|X-\mu| \geq t) = 0, \ \forall t > 0$

Definition 3.6 (characteristic function, TBp. 161)

The **characteristic function (chf)** of a random variable X is

mgf (LNp.2-46) cf.

Fourier transformation

$\because e^{iu} = \cos(u) + i\sin(u)$

$$\phi_X(t) = E(e^{itX}) = E[\cos(tX)] + i \cdot E[\sin(tX)], = \begin{cases} \int e^{itx} f_X(x)\,dx \\ \sum_x e^{itx} p_X(x) \end{cases}$$

where $i = \sqrt{-1}$, and the joint characteristic function of $X_1, X_2, \ldots, X_n$ is

$$\phi_{X_1X_2\cdots X_n}(t_1, t_2, \ldots, t_n) = E(e^{it_1X_1+it_2X_2+\cdots+it_nX_n}).$$

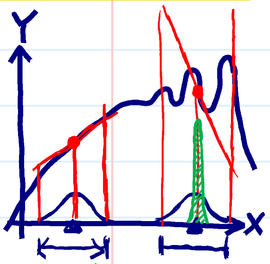
cf. joint mgf (LNp.2-48)

Theorem 3.7 (properties of characteristic function)

mgf might not exist ($\because E(e^{tX}) = \infty$) cf.

1. The characteristic function always exists. $\because E|e^{itX}| = E(1) = 1 < \infty$
2. If $M_X(t)$ exists, then $\phi_X(t) = M_X(it)$.
3. uniqueness theorem
4. (FYI) inversion theorem:
   - discrete case: $p_X(x) = \lim_{T\to\infty} \int_{-T}^{T} e^{-itx}\phi_X(t)\,dt$
   - continuous case: $f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_X(t)\,dt$
5. The properties of characteristic function are similar to those of moment generating function.

- **δ method**



What if Var(X) very small?

Question 3.3

Recall Chebyshev's inequality (LNp.43)

Let $Y = g(X)$. Suppose we only know the mean μ_X and variance σ_X^2 of X, but **not** the entire distribution (i.e., do not know cdf, pdf/pmf of X). Can we derive the distribution of Y? If not, can we "roughly" describe the mean and variance of Y? (**Note.** $E[g(X)] \neq g[E(X)]$.)

Theorem 3.9 (δ method for univariate case, TBp. 162)

Thm 3.1 (LNp.41)

When does it hold?

$$Y = g(X) \boxed{\approx} g(\mu_X) + (X-\mu_X)g'(\mu_X) \quad \text{(by Taylor expansion)} \quad \cancel{+\cdots}$$

$$\Rightarrow \quad E[g(X)] \approx g(\mu_X) \qquad \text{cf.}$$

$$Var[g(X)] \approx Var(X)[g'(\mu_X)]^2$$

or
$$Y = g(X) \boxed{\approx} g(\mu_X) + (X-\mu_X)g'(\mu_X) + \frac{1}{2}(X-\mu_X)^2 g''(\mu_X) \quad \cancel{+\cdots}$$

$$\Rightarrow \quad E[g(X)] \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$$

Note. How good these approximations are depends on whether g can be reasonably well approximated by the 1st or 2nd order polynomials in a neighborhood of μ_X and on the size of σ_X. ← check Chebyshev's inequality (LNp.43)

Theorem 3.10 (δ method for multivariate case, TBp. 165)

Suppose that we only know μ_X, μ_Y, σ_X^2, σ_Y^2, σ_{XY}, but not the joint dist. of X, Y

Function of two univariate random variables $Z = g(X,Y)$:
Let $\mu = (\mu_X, \mu_Y)$.

$$Z = g(X,Y) \approx g(\mu) + (X-\mu_X)\frac{\partial g(\mu)}{\partial x} + (Y-\mu_Y)\frac{\partial g(\mu)}{\partial y} \quad \cancel{+\cdots}$$

$$\Rightarrow \quad \mathrm{E}(Z) \approx g(\mu)$$

$$\mathrm{Var}(Z) \approx \sigma_X^2\left[\frac{\partial g(\mu)}{\partial x}\right]^2 + \sigma_Y^2\left[\frac{\partial g(\mu)}{\partial y}\right]^2 + 2\sigma_{XY}\left[\frac{\partial g(\mu)}{\partial x}\right]\left[\frac{\partial g(\mu)}{\partial y}\right]$$

or
$$g(X,Y) \approx g(\mu) + (X-\mu_X)\frac{\partial g(\mu)}{\partial x} + (Y-\mu_Y)\frac{\partial g(\mu)}{\partial y} + \frac{1}{2}(X-\mu_X)^2\frac{\partial^2 g(\mu)}{\partial x^2} + (X-\mu_X)(Y-\mu_Y)\frac{\partial^2 g(\mu)}{\partial x \partial y} + \frac{1}{2}(Y-\mu_Y)^2\frac{\partial^2 g(\mu)}{\partial y^2} \quad \cancel{+\cdots}$$

$$\Rightarrow \quad E[g(X,Y)] \approx g(\mu) + \frac{1}{2}\sigma_X^2\frac{\partial^2 g(\mu)}{\partial x^2} + \sigma_{XY}\frac{\partial^2 g(\mu)}{\partial x \partial y} + \frac{1}{2}\sigma_Y^2\frac{\partial^2 g(\mu)}{\partial y^2}$$

Note. The general case of a function of n random variables can be worked out similarly.

- **Reading**: textbook, Chapter 4 ← required
- **Further Reading**: Roussas, 5.1, 5.3, 5.4, 5.5, 6.1, 6.2, 6.4, 6.5 ← optional

Definition 4.13 (Normal distribution $N(\mu, \sigma^2)$, sec. 2.2.3)

Shape

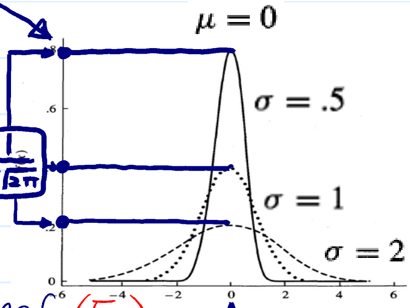
- **pdf:** $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$

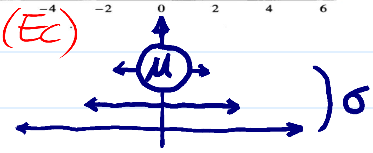
a pdf? (Ec) ← use polar coordinates

- **mgf:** $e^{\mu t + \frac{\sigma^2 t^2}{2}}, \ t \in \mathbb{R}$ ← use STO (Ec)
- **mean:** μ ← show that its pdf is symmetric about μ. (Ec) / use mgf
- **variance:** σ^2 ← Find $E(X^2)$ using mgf (Ec)
- **parameter:** $\mu \in \mathbb{R}, \sigma > 0$

meaning of parameters

$\frac{1}{\sigma\sqrt{2\pi}}$





Notes:

1. bell-shaped pdf (symmetric about μ, where it has maximum, and falls off in the rate determined by σ)
2. play a central role in probability and statistics (e.g., CLT, Chapter 5)
3. $X \sim N(\mu, \sigma^2) \Rightarrow$ for $a, b \in \mathbb{R}$, $aX + b \sim N(a\mu + b, a^2\sigma^2)$. In particular, $\frac{X-u}{\sigma} \sim N(0, 1)$. ← prove using mgf (Ec)

   standari-zation

4. Let $X_1, \ldots, X_k$ be independent, and $X_i \sim N(\mu_i, \sigma_i^2)$. Then $Y = X_1 + \cdots + X_k \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2)$ ← prove using mgf (Ec)
5. by 3 and 4, let $X_1, \ldots, X_k$ be i.i.d $\sim N(\mu, \sigma^2)$, then $\overline{X}_k \sim N(\mu, \frac{\sigma^2}{k})$.

   $= \frac{X_1 + \cdots + X_k}{k}$ $\xrightarrow{k \to \infty} 0$