Formulary

10. Januar 2020

Discrete distributions:

Definition 1. (Uniform): A random variable X taking values in the finite set $\{x_1, ..., x_n\}$ for some integer $n \geq 1$ is said to be uniform if $P(X = x_i) = \frac{1}{n} \forall i \in \{1, ..., n\}$. We write $X \sim \mathcal{U}\{x_1, ..., x_n\}$.

Definition 2. (Bernoulli): $X_1, ..., X_n : \Omega \to \{0, 1\}$ are usually referred to as Bernoulli random variables or Bernoulli trials with ßuccessprobability p. We write $X_1, ..., X_n \sim^{\text{i.i.d.}} \text{Ber}(p)$.

Definition 3. (Binomial): Let $X_1, ..., X_n$ be i.i.d. random variables such that $P(X_i = 1) = p = 1 - P(X_i = 0) \forall \in \{1, ..., n\}$ and some fixed $p \in (0, 1)$. Define $S_n := X_1 + \cdots + X_n$ called a Binomial random variable with parameters n and p. We write $S_n \sim \text{Bin}(n, p)$.

Proposition 1. Let $S_n \sim \text{Bin}(n, p)$ ($n \in \mathbb{N}$, $p \in (0, 1)$). Then:
a) $P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for $k \in \{0, 1, ..., n\}$)

Proposition 2. $X_i \sim \text{Bin}(n_i, p) \Rightarrow S_n := \Sigma Xi := X_1 + ... + X_k \sim \text{Bin}(\Sigma n_i, p)$ [L13]

Proposition 3. (Geometric): Let $(X_n)_{n \geq 1}$ be i.i.d. Bernoulli random variables with success probability $p \in (0, 1)$. Define the random variable X as the first integer $i \geq 1$ for which $X_i = 1$. X is called a geometric random variable with success probability p. We write $X \sim \text{Geo}(p)$ or $X \sim \mathcal{G}(p)$.

Definition 4. (Negative Binomial): Let $(X_n)_{n \geq 1}$ be i.i.d. Bernoulli random variables with success probability $p \in (0, 1)$. Define X as the random variable which is equal to the first integer i for which $\sum_{j=1}^{i} X_j = r$ where r is some fixed integer $\in \{1, 2, ...\}$. X is said to be a Negative Binomial random variable with parameters p and r. Note that by definition $X \geq r$. We write $X \sim \text{NB}(r, p)$.

Proposition 4.

Proposition 5. $X_i \sim NB(r_i, p) \Rightarrow \Sigma X_i := X_1 + ... + X_k \sim NB(\Sigma r_i, p)$ [L13]

Definition 5. (Hypergeometric): Consider a population with N distinct individuals and composed exactly of D individuals of type I and $N - D$ individuals of type II. Draw from this population n individuals at random and without replacement (an individual cannot be selected more than once). Define $X =$ number of individuals of type I among the n selected ones. Then, X is called a Hypergeometric random variable with parameters n, D and N. We write $X \sim \text{Hypergeo}(n, D, N)$.

Definition 6. (Poisson): A random variable X taking its values in $\mathbb{N}_0$ is called Poisson random variable if $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for all $k \in \mathbb{N}_0$ and some given $\lambda \in (0, \infty)$. We write $X \sim \text{Poi}(\lambda)$ or $X \sim \mathcal{P}(\lambda)$. Here, λ is called rate or intensity.

Theorem 1. Let $\lambda \in (0, \infty)$ and consider $S_n \sim \text{Bin}(n, \frac{\lambda}{n})$ for $n \geq [\lambda + 1]$. Then for $k \in \mathbb{N}_0$

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

Proposition 6. $Xi \sim Poi(\lambda_i) \Rightarrow \Sigma X_i := X_1 + ... + X_k \sim Poi(\Sigma\lambda_i)$ [L13]

Continuous distributions [L14]:

Definition 7. (Uniform): If $X \sim f$ with $f(x) := 1_{[0,1]}(x)$ then X is said to be a uniform random variable on [0,1]. We write $X \sim U([0, 1])$.

Definition 8. (Beta): If $X \sim f$ with $f(x) := \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{(0,1)}(x)$ with $\alpha, \beta > 0$ and $\Gamma(a) := \int_0^\infty t^{\alpha-1}e^{-t}dt$ for $a \in (0, \infty)$ then X is said to be a Beta random variable with parameters α and β.

Definition 9. (Exponential): If $X \sim f$ with $f(x) := \lambda e^{-\lambda x} 1_{(0,\infty)}(x)$, then X is said to be an exponential random variable with parameters/rate/intensity lambda. We write $X \sim Exp(\lambda)$.

Definition 10. (Gamma): If $X \sim f$ with $f(x) := \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x} 1_{(0,\infty)}(x)$ for $alpha, beta > 0$ then X is said to be a gamma random variable with shape parameter alpha and rate/intensity beta. We write $X \sim G(\alpha, \beta)$. Note that if $alpha = 1$, then $G(\alpha, \beta) =^d \text{Exp}(\beta)$.

Exercise 1. (S7E3d)) Let $X_1, ..., X_n \sim^{\text{i.i.d.}} G(\alpha_i, \beta),\ \alpha_1, ..., \alpha_n,\ \beta > 0$, then $X_1 + \cdots + X_n \sim G(\alpha_1 + \cdots + \alpha_n, \beta)$.

Definition 11. (Normal/Gaussian): If $X \sim f$ with $f(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in R$ then X is said to be a standard Normal Gaussian random variable. We write $X \sim N(0,1)$. Note that here 0 and 1 correspond to expectation and variance of X respectively.

Definition 12. A usual notation for $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is ϕ and for its cdf $x \mapsto \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ is Φ.

Proposition 7. Let $X_1, ..., X_n \sim^{i.i.d.} N(0,1)$ then $X_1 + \cdots + X_n \sim N(0,n)$.

Theorem 2. Let X be some application g defined on an open set $O \subset \mathbb{R}$ s.t. $g \in C^1(O)$, g is strictly monotone on O with $g'(x) \neq 0 \forall x \in O$ and $P(X \in O) = 1$.
Then the random variable $Y = g(X)$ is absolutely continuous with density equal a.e. to

$$f_Y(y) = \frac{f_X \circ g^{-1}(y)}{|g' \circ g^{-1}(y)|} 1_{y \in g(O)}(y).$$

Definition 13. (Normal/Gaussian): A random variable X is said to have normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$ if it has density $f(x) = f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in R$. We then write $X \sim N(\mu, \sigma^2)$.

Theorem 3. a) If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
$Z \sim N(0,1)$ then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$
b) If $X \sim N(\mu, \sigma^2)$ then $\forall x \in \mathbb{R} : F_X(x) = \Phi(\frac{x-\mu}{\sigma^2})$. In particular, $F_X(\mu) = \frac{1}{2}$.
c),d)

Theorem 4. Let $X_1, ..., X_n$ be $n \geq 2$ independent random variables such that $X_i \sim N(\mu_i, \sigma_i^2)$, $\mu_i \in \mathbb{R}$ and $\sigma_i \in (0, \infty)$. Let $S_n := \sum_{i=1}^n X_i$.
Then $S_n \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

Definition 14. (S8E4; Quantile transformation) Given a cdf F, the **quantile** t_α of order $\alpha \in (0,1)$ is defined as

$$t_\alpha = \inf\{t : F(t) \geq \alpha\} =: F^{-1}(\alpha)$$

F^{-1} denotes the generalized inverse of F. When the latter is bijective (at least in the neighborhood of t_α), then F^{-1} is the inverse of F in the classical sense.

Definition 15. Let $(Z_n)_{n\geq 1}$ be a sequence of random variables (not necessarily defined on the same probability space). We say that $(Z_n)_{n\geq 1}$ **converges in distribution** (in law or weakly) towards $Z \sim N(0,1)$ if $F_{Z_n}(x) \to_{n\to\infty} F(x) = \Phi(x) \forall x \in \mathbb{R}$.
We write $Z_n \to^d_{n\to\infty} Z$ or ....

Theorem 5 (Central Limit Theorem)**.** Let $X_1, ..., X_n$ be i.i.d. random variables with expectation μ and variance σ ($\mu \in \mathbb{R}, \sigma \in (0,\infty)$). Then

$$Z_n = \frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \to^d_{n\to\infty} Z \sim N(0,1)$$

(recall $\bar{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$) or equivalently $\frac{S_n - n\mu}{\sqrt{n}\sigma} \to^d_{n\to\infty} Z \sim N(0,1)$.

Theorem 6 (de Moivre-Laplace Theorem)**.** Let $X_n \sim \mathrm{Bin}(n,p)$ $p \in (0,1)$. Then

$$\frac{X_n - np}{\sqrt{np(1-p)}} \to^d_{n\to\infty} Z \sim N(0,1)$$

$$\iff P(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x) \to_{n\to\infty} \Phi(x) \forall x \in \mathbb{R}.$$

Theorem 7 (Iterated Expectation Formula)**.** Provided $E[X]$ exists:

$$E[X] = E[E[X|Y]].$$

Theorem 8 (Iterated Variance Formula)**.** Provided $Var(X)$ exists:

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y]).$$

Definition 16. Let (X,Y) be a random pair s.t. $0 < \mathrm{Var}(Y) < \infty$. The **correlation** between X and Y is defined as $\mathcal{C}_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$

Theorem 9.

1. $\forall c \in \mathbb{R}, X : E[X] < \infty \implies \mathrm{Cov}(X,c) = 0$,
2. Provided $Var(X) < \infty$, $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$,
3. Provided $\forall i,j : \mathrm{Cov}(X_i, Y_j) < \infty$ it holds that

$$\mathrm{Cov}(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \,\mathrm{Cov}(X_i, Y_j),$$

4. If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0 = \rho_{X,Y}$

5. $|\rho_{X,Y}| \leq 1$. Furthermore,

$$\rho_{X,Y} = 1 \iff \exists a \in \mathbb{R}, b > 0 : P(Y = a + bX) = 1,$$
$$\rho_{X,Y} = -1 \iff \exists a \in \mathbb{R}, b < 0 : P(Y = a + bX) = 1.$$

Definition 17. Let (X, Y) be a discrete random pair with joint pmf p. In this case, for any function g on $\mathbb{R}^2$ $Z = g(X, Y)$ is a discrete random variable. We call its (marginal) pmf the function

$$p_z(z) = P(Z = z) = \sum_{(x,y):g(x,y)=z} p(x, y).$$

Proposition 8. Let (X, Y) be a discrete random pair. Then:

1. $E[g(Y)|Y = y] = g(y)$ for any function g on $\mathbb{R}^2$)
2. $E[Xg(Y)|Y = y] = g(y)E[X|Y = y]$ (provided that $E[X|Y = y]$ exists) for g on $\mathbb{R}^2$.

Proposition 9. Let (X, Y) be a discrete random pair. Then X and Y are independent if and only if.

1. $\forall x \in \mathbb{R}^2, y \in \mathbb{R} : p_Y(y) > 0 \implies p(x|y) = p_X(x)$, or
2. $\forall y \in \mathbb{R}^2, x \in \mathbb{R} : p_X(x) > 0 \implies p(y|x) = p_Y(y)$

Proposition 10. If (X, Y) is a discrete random vector s.t. X and Y are independent , then for any function g on $\mathbb{R}$ s.t. $E[g(X)]$ exists, we have $E[X|Y = y] = E[g(X)]$.

Definition 18. 1. Let (X, Y) be a discrete random pair with joint pmf p. Denote $E[X|Y = y] = \mu_X(y)$. Then, the conditional variance of X given $Y = y$ is

$$var(X|Y = y) = \sum_x p(x|y)(x - \mu_x(y))^2.$$

2. For $y \in Y = \{y_1, ...\}$ it is clear that $E[X|Y = y]$ depends on............

Definition 19. A random vector $X = (X_1, ..., X_n)^T$ is said to be absolutely continuous with respect to Lebesgue measure λ_n defined on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ if its induced probability measure $P_X << \lambda_n$. In this case, it follows from the Radon-Nikodyn theorem that there exists a non-negative measurable function f defined on $\mathbb{R}^n$:

$$P_X(B) = \int_B f(x_1, ..., x_n) d\lambda_n(x_1, ..., x_n) =: \int_B f(x_1, ..., x_n) dx_1, ..., dx_n.$$

f is called the (joint) density of X.

Definition 20. Let $(X, Y)^T$ be the 2-d random vector with density f (on $\mathbb{R}^2$).

1. The conditional density of X given $Y = y$ is defined as $f(x|y) = \frac{f(x,y)}{f_Y(y)}$ wherever $f_Y(y) > 0$ and $f(x|y) = 0$ wherever $f_Y(y) = 0$.
2. $f(y|x)$ is defined analogously.

Definition 21. (conditional variance). Conditional variance is defined as $\mathrm{Var}(X|Y) = h(Y)$, where $h(y) := \int_{\mathbb{R}} (y - E(y|X = x))^2 f(y|x) dy$.

Remark 1. It now still holds that:

1. $E[X] = E[E[X|Y]]$ (whenever $E[X] < \infty$) and
2. $\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$ (whenever $\mathrm{Var}(X) < \infty$).

Theorem 10. Let $X = (X_1, ..., X_n)^T$ be an absolutely continuous random vector with joint density $f = f_X$. Consider $g : \mathcal{O} \to g(\mathcal{O}) \subset \mathbb{R}^n$ a bijective application defined on an open set $\mathcal{O} \subset \mathbb{R}^n$ s.t. $g \in C^1(\mathcal{O})$ and

$$J_g(x) = \det \begin{pmatrix} \partial_1 g_1 & \cdots & \partial_n g_1 \\ \vdots & & \vdots \\ \partial_1 g_n & \cdots & \partial_n g_n \end{pmatrix}.$$

Assume that $P(X \in \mathcal{O}) = 1$. Then $Y = g(X)$ is an absolutely continuous with joint density

$$f_Y(y) = \frac{f_X \circ g^{-1}(y)}{|J_g \circ g^{-1}(y)|} 1_{g(\mathcal{O})}.$$

Remark 2. Recall that $X = (X_1, ..., X_n)^T \sim N(\mu, \Sigma)$, that is X is a Gaussian vector with expectation $\mu = (E(X_1), ..., E(X_n))^T$ and covariance matrix Σ ($\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$). If $X =^d \mu + AZ$, where A is a square root of Σ (i.e. $\Sigma = AA^T$) and $Z = (Z_1, ..., Z_n)^T \sim N(0, id_n)$ meaning that $Z_1, ..., Z_n \sim^{\text{i.i.d.}} N(0, 1)$ and

$$f_Z(z) = \prod_{i=1}^n f_{Z_i}(z_i) = \frac{1}{(2\pi)^{n/2}} e^{\frac{1}{2}\sum_{i=1}^n z_i^2} = \frac{1}{(2\pi)^{n/2}} e^{\frac{1}{2} z^T z}$$

for $z = (z_1, ..., z_n)^T \in \mathbb{R}^n$.

Theorem 11. If $X \sim N(\mu, \Sigma)$ and Σ is invertible (pos. definite), then X is absolutely continuous with density

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \ \forall X = (X_1, ..., X_n)^T \in \mathbb{R}.$$

- $X_1, ..., X_n$ independent with induced probability measure $P_{X_i}(B) = P_{X_1}(B)\ \forall B \in \mathcal{B}(\mathcal{X}) = \sigma-$Algebra on $\mathcal{X}$.

- We will assume that $P_{X_1} \in \mathcal{P}$, where $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^d$ is the parameter space and P_θ a probability measure on $\mathcal{X}$. Then $\mathcal{P}$ is called a parametric model.

- We further assume that $\forall \theta \in \Theta : P_\theta$ admits a density w.r.t. some σ-finite dominating measure μ on $(\mathcal{X}, \mathcal{B})$. We will denote this densit by $f(\cdot|\theta)$.

Definition 22. Let $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$ with $\theta_0 \in \Theta \subset \mathbb{R}^d,\ d \geq 1$. $\hat{\theta}$ is an **estimator** for θ_0, based on $X_1, ..., X_n$, if $\hat{\theta}$ is a **statistic** of $X_1, ..., X_n$, that is any quantity of the form $T(X_1, ..., X_n)$, where T is a measurable map on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Definition 23. Let $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$ with $\theta_0 = (\theta_{01}, ..., \theta_{0d})^T \in \Theta \subset \mathbb{R}^d,\ d \geq 1$. Also suppose that if $X \sim f(\cdot|\theta)$ the moments $E[X], ..., E[X^d]$ exist and

$$\begin{aligned}
\theta_1 &= \Psi_1(E[X], ..., E[X^d]) \\
\theta_2 &= \Psi_2(E[X], ..., E[X^d]) \\
&\vdots \\
\theta_d &= \Psi_d(E[X], ..., E[X^d])
\end{aligned}$$

with $\Psi_1, ..., \Psi_d$ some measurable functions.
The **moment estimator**, $\hat{\theta}$, of θ_0 is obtained by replacing $E[X^j]$ by its empirical estimator: $\frac{1}{n}\sum_{i=1}^n X_i^j$ for $j \in \{1, ..., d\}$, i.e.

$$\begin{aligned}
\theta_1 &= \Psi_1(\frac{1}{n}\sum_{i=1}^n X_i, ..., \frac{1}{n}\sum_{i=1}^n X_i^d) \\
&\vdots \\
\theta_d &= \Psi_d \frac{1}{n}\sum_{i=1}^n X_i, ..., \frac{1}{n}\sum_{i=1}^n X_i^d).
\end{aligned}$$

Definition 24.

- Let $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$ with some unknown $\theta_0 \in \Theta$, the **likelihood function** defined by Θ is given by $L(\theta) = \prod_{i=1}^n f(x_i|\theta),\ \theta \in \Theta$.

- The **maximum likelihood estimator** (MLE) is defined by $\hat{\theta} = \text{argmax}_{\theta\in\Theta} L(\theta)$. Provided it exists and is unique, and is a measurable map of $X_1, ..., X_n$).

Now let $d = 1$, i.e. $\Theta \subset \mathbb{R}$.

Theorem 12. Let $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$ with $\theta_0 \in \Theta$. Let $\hat{\theta}$ $(= \hat{\theta}_n)$ be the MLE based on $X_1, ..., X_n$. Under some regularity conditions (on $(x, \theta) \mapsto f(x, \theta)$ the MLE $\hat{\theta}$ is consistent, that is $\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$ $\forall\epsilon > 0$ ($\iff$ $\hat{\theta} \to^P \theta_0$ as $n \to \infty$).

Theorem 13. Let $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$ with $\theta_0 \in \Theta$. Let $\hat{\theta}$ $(\hat{\theta}_n)$ be the MLE based on $X_1, ..., X_n$ under additional regularity conditions, it holds that

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, \frac{1}{I(\theta_0)},$$

where

$$I(\theta_0) = E[(\frac{\partial \log(f(x_1, \theta))}{\partial\theta}|_{\theta=\theta_0})^2] = -E[\frac{\partial^2 \log(f(x_1, \theta))}{\partial\theta^2}|_{\theta=\theta_0}]$$

is called the Fischer information of the model for $\theta = \theta_0$ and is assumed to belong to $(0, \infty)$.

Suppose we have two estimators $\hat{\theta} = \hat{\theta_n}$ and $\tilde{\theta} = \tilde{\theta_n}$ computed on the basis of i.i.d. random variables $X_1, ..., X_n \sim f(\cdot|\theta_0)$, $\theta_0 \in \Theta$.

Definition 25.
- An estimator $\hat{\theta_n}$ of $\theta_0 \in \Theta$ computed on the basis of i.i.d. random variables $X_1, ..., X_n \sim f(\cdot|\theta_0)$ is said to be **unbiased** if $E[\hat{\theta_n}] = \theta_0$ ($\forall\theta_0$ `unknown in` Θ)
- The **mean square error** of $\hat{\theta_n}$ is defined as $\text{MSE}(\hat{\theta_n}) := E[(\hat{\theta_n} - \theta_0)^2]$ (provided that $E[\hat{\theta_n}^2] < \infty$).

Remark 3. $\text{MSE}(\hat{\theta}_n) = var(\hat{\theta}) + (E[\hat{\theta}_n] - \theta_0)^2$ and $bias(\hat{\theta}_n) := E[\hat{\theta}_n] - \theta_0$. If $\hat{\theta}_n$ is unbiased i.e. $bias(\hat{\theta}_n) = 0$, then $\text{MSE}(\hat{\theta}_n) = \text{Var}(\hat{\theta}_n)$.

Definition 26. We say that $\hat{\theta}_n$ is **more efficient** than $\tilde{\theta}_n$ if $\text{MSE}(\hat{\theta}_n) \leq \text{MSE}(\tilde{\theta}_n)$. We define the **efficiency** of $\hat{\theta}_n$ relative to $\tilde{\theta}_n$ as $eff(\hat{\theta}_n, \tilde{\theta}_n) := \frac{\text{MSE}(\tilde{\theta})}{\text{MSE}(\hat{\theta}_n})$. (So if $eff(\hat{\theta}_n, \tilde{\theta}_n) \geq 1$ then $\hat{\theta}_n$ is more efficient than $\tilde{\theta}_n$)

Theorem 14. Let $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$ for some $\theta_0 \in \Theta$. If $\hat{\theta}_n = T(X_1, .., X_n)$ with T some measurable map is an unbiased estimator of θ_0, then $\text{Var}(\hat{\theta}_n) \geq \frac{1}{nI(\theta_0}$, where $I(\theta_0)$ is the Fischer Information (under some regularity conditions).

Definition 27. A statistic $T(X_1, ..., X_n)$ is said to be **sufficient** for θ if the conditional distribution of $(x_1, ..., x_n)^T$ given $T(X_1, ..., X_n) = t$ does not depend on θ, whenever $X_1, ..., X_n \sim^{\text{i.i.d.}} f(\cdot|\theta_0)$.

Theorem 15. (Factorization Theorem) A statistic $T(X_1, ..., X_n)$ is sufficient for θ if and only if there exist non-negative functions g and h s.t.

$$L(\theta) = \Pi_{i=1}^n f(X_i|\theta) = g(T(X_1, ..., X_n), \theta) \cdot h(X_1, ..., X_n)$$

Corollary 1. (Sh 12 Ex 2.b)) If T is sufficient for θ then cT is sufficient for θ for any $c \in \mathbb{R}^\times$.