Performance Evaluation and Networks

Statistics II

Estimators

Framework: statistical inference from a sample $(x_1, \dots, x_n) \in E^n$ seen as a realization of a random vector $(X_1, \dots, X_n)$ following a law with an unknown parameter $\theta \in \Theta$ (usually $\Theta = \mathbb{R}$ or $\mathbb{R}^d$). This parameter may fully characterize the law or may be just one parameter among others.

Definition (Estimator of a parameter)

- *function $\theta_n : E^n \to \Theta$ the parameter set, for size n samples*
- *family of functions $(\theta_n)_{n \in \mathbb{N}^*}$ to deal with any sample size*

Examples: $\frac{1}{n}\sum_1^n x_i$, $\max_{1 \le i \le n} x_i$, constant function c, ...

Point estimation: find some estimators such that the random variable $\widehat{\theta}_n \overset{\text{def}}{=} \theta_n(X_1, \dots, X_n)$ gives some information about θ with high probability, so that there is a high probability that the value $\theta_n(x_1, \dots, x_n)$ from the sample holds some information about θ.

Classical properties

Definition (estimator properties)

Let $(\theta_n)_{n\in\mathbb{N}^*}$ a family of estimators for some parameter $\theta \in \mathbb{R}$ (or $\mathbb{R}^d$) of the random vector $(X_n)_{n\in\mathbb{N}^*}$, and denote $\widehat{\theta}_n = \theta_n(X_1,\ldots,X_n)$. Such estimators are called:

- ***unbiased*** *if* $\forall n \in \mathbb{N}^*, \mathbb{E}(\widehat{\theta}_n) = \theta$
- ***asymptotically unbiased*** *if* $\lim_{n\to\infty} \mathbb{E}(\widehat{\theta}_n) = \theta$
- ***consistent*** *if* $\widehat{\theta}_n \xrightarrow{P} \theta$ *(convergence in probability)*
- ***strongly consistent*** *if* $\widehat{\theta}_n \xrightarrow{a.s.} \theta$ *(almost sure convergence)*

Example: let $(X_n)_{n\geq 1}$ i.i.d. random variables of finite mean μ and consider $\overline{\mu}_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_1^n x_i$, then $\overline{\mu}_n$ is an estimator of μ which is *unbiased* (linearity of $\mathbb{E}$) and *strongly consistent* (strong law of large numbers).

Classical properties

Interpretation of estimator properties

- **unbiased** if $\forall n \in \mathbb{N}^*, \mathbb{E}(\widehat{\theta}_n) = \theta$
  Given size n, the estimator may fail once (the outcome ω may yield $\widehat{\theta}_n(\omega) = \theta_n(x_1, \dots, x_n) \neq \theta$) but generating many samples will give the right value θ on average.
- **asymptotically unbiased** if $\lim_{n\to\infty} \mathbb{E}(\widehat{\theta}_n) = \theta$
  Same as above if sample size also grows.
- **consistent** if $\widehat{\theta}_n \xrightarrow{P} \theta$ (convergence in probability)
  Larger samples have a higher proportion of good estimates. But if one makes a single sample grow, the estimator may fail in a recurrent way.
- **strongly consistent** if $\widehat{\theta}_n \xrightarrow{a.s.} \theta$ (almost sure convergence)
  The larger the sample, the better the estimation.

Designing estimators

Some approaches:

- Method of moments
- Maximum likelihood estimation (MLE)
- Maximum spacing estimation (MSE)

Method of moments

Idea: match empirical moments from the model $(X_1, \ldots, X_n)$ and from the data $(x_1, \ldots, x_n)$, then solve for unknown parameters.

General scheme for models with d parameters $\theta_1, \ldots, \theta_d$

1. Express empirical moments for the model: $\overline{m}_k = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i^k)$ as function of $\theta_1, \ldots, \theta_d$ (if you can)
2. Consider empirical moments for the sample: $m_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k$
3. Choose some values of k for which you match those moments: $\overline{m}_k = m_k$
4. Solve this system of equations for unknown $\theta_1, \ldots, \theta_d$

Advice: best suited for models $(X_1, \ldots, X_n)$ with i.i.d. random variables, where it often yields consistent estimators (law of large numbers).

Method of moments

Example: $(X_1, \dots, X_n)$ i.i.d. $\sim \mathscr{U}([a,b])$ uniform over $[a,b]$, where $\overline{m}_1 = \frac{a+b}{2}$ and $\overline{m}_2 = \frac{b^3-a^3}{2(b-a)}$. Consider the system of equations $\overline{m}_1 = m_1$ and $\overline{m}_2 = m_2$, its solution is $a_n = m_1 - \sqrt{3(m_2 - m_1^2)}$ and $b_n = m_1 + \sqrt{3(m_2 - m_1^2)}$.

Example: $(X_1, \dots, X_n)$ i.i.d. $\sim \mathscr{B}(p)$ Bernoulli, where $\overline{m}_1 = p$ and $\overline{m}_2 = p$. Then depending on the choice of equation, one get the estimators $p_n = m_1$ or m_2, which are the same for these particular laws.

Maximum likelihood estimation (MLE)

Definition (Likelihood of a sample)

- *Let $\theta \in \mathbb{R}$ and f_θ the law of vector $(X_1, \ldots, X_n)$ which supposedly generated the sample $(x_1, \ldots, x_n)$, the **likelihood** of $(x_1, \ldots, x_n)$, denoted by $L_n(\theta)$ or $L_n(\theta, x_1, \ldots, x_n)$, is $f_\theta(x_1, \ldots, x_n)$.*
- *Particular case (i.i.d. sampling): when $X_1, \ldots, X_n$ are i.i.d. of law f_θ, then $L_n(\theta) = f_\theta(x_1) \cdots f_\theta(x_n)$.*

Definition (Maximum Likelihood Estimator)

*An estimator θ_n of θ is called a **maximum likelihood estimator** if θ_n maximizes $L_n(\theta)$, i.e. $\theta_n(x_1, \ldots, x_n) = \underset{\theta \in \mathbb{R}}{\operatorname{argmax}}\, L_n(\theta, x_1, \ldots, x_n)$*

Maximum likelihood estimation (MLE)

In practice: maximize the function or its logarithm, thus study $\frac{\partial L_n(\theta)}{\partial \theta}(x_1,\dots,x_n)$ or $\frac{\partial \log L_n(\theta)}{\partial \theta}(x_1,\dots,x_n)$, when defined, to find θ_n.

Example: faulty machine with i.i.d. Bernoulli errors $\mathscr{B}(p)$. MLE of p for sample $(x_1,\dots,x_n) \in \{0,1\}^n$?

Example: Poisson traffic with i.i.d. exponential inter-arrivals of parameter λ. MLE of λ for sample $(x_1,\dots,x_n) \in \mathbb{R}_+^n$?

Maximum likelihood estimation (MLE)

In practice: maximize the function or its logarithm, thus study $\frac{\partial L_n(\theta)}{\partial \theta}(x_1, \ldots, x_n)$ or $\frac{\partial \log L_n(\theta)}{\partial \theta}(x_1, \ldots, x_n)$, when defined, to find θ_n.

Example: faulty machine with i.i.d. Bernoulli errors $\mathscr{B}(p)$. MLE of p for sample $(x_1, \ldots, x_n) \in \{0,1\}^n$?

Let $n_0 = |\{i \mid x_i = 0\}|$ and $n_1 = |\{i \mid x_i = 1\}|$. Study the variations of $p \mapsto L_n(p, x_1, \ldots, x_n) = p^{n_1}(1-p)^{n_0}$ by differentiating.
The maximum is reached for $p_n = n_1 / (n_0 + n_1) = \sum_{i=1}^{n} x_i / n$

Example: Poisson traffic with i.i.d. exponential inter-arrivals of parameter λ. MLE of λ for sample $(x_1, \ldots, x_n) \in \mathbb{R}_+^n$?

Maximum likelihood estimation (MLE)

In practice: maximize the function or its logarithm, thus study $\frac{\partial L_n(\theta)}{\partial\theta}(x_1,\dots,x_n)$ or $\frac{\partial \log L_n(\theta)}{\partial\theta}(x_1,\dots,x_n)$, when defined, to find θ_n.

Example: faulty machine with i.i.d. Bernoulli errors $\mathscr{B}(p)$. MLE of p for sample $(x_1,\dots,x_n)\in\{0,1\}^n$?
Let $n_0 = |\{i \mid x_i = 0\}|$ and $n_1 = |\{i \mid x_i = 1\}|$. Study the variations of $p \mapsto L_n(p, x_1,\dots,x_n) = p^{n_1}(1-p)^{n_0}$ by differentiating.
The maximum is reached for $p_n = n_1/(n_0+n_1) = \sum_{i=1}^n x_i/n$

Example: Poisson traffic with i.i.d. exponential inter-arrivals of parameter λ. MLE of λ for sample $(x_1,\dots,x_n)\in\mathbb{R}_+^n$?
Study the variations of $\lambda \mapsto L_n(\lambda, x_1,\dots,x_n) = \lambda^n e^{-\lambda(x_1+\cdots+x_n)}$ by differentiating.
The maximum is reached for $\lambda_n = n/(x_1+\cdots+x_n)$

Consistency of MLE

Hypotheses: model = random vector X following a law from the family $f_\theta, \theta \in \Theta$, and θ_n some MLE estimator of θ.

Theorem (Consistency of MLE)

If the next conditions are satisfied:

- *identification:* $\theta_1 \neq \theta_2 \Rightarrow f_{\theta_1} \neq f_{\theta_2}$
- *compactness:* Θ *is compact*
- *continuity:* $(\theta, x) \longmapsto f_\theta(x)$ *is continuous*
- *bounded entropy:* $\forall \theta \in \Theta$, $H_\theta = -\int f_\theta(x) \log f_\theta(x) dx < +\infty$

Then $\widehat{\theta}_n \xrightarrow{P} \theta$

Comparison of estimators

Some quality index: mean squared error (MSE)

$$R(\theta_n, \theta) \stackrel{\text{def}}{=} \mathbb{E}(\widehat{\theta}_n - \theta)^2$$

Definition (Dominant estimators)

Let θ_n and ψ_n two estimators of θ, θ_n is said to dominate ψ_n if $\forall \theta$, $R(\theta_n, \theta) \leq R(\psi_n, \theta)$ with strict inequality somewhere.

Remark: there is not always an estimator dominating all others.

Fisher Information

Hypotheses: model = random vector X following a law from the family $f_\theta, \theta \in \Theta$.

Definition (Fisher Information)

If the next conditions are satisfied:

- *the support of f_θ is indep of θ*
- $\frac{\partial f_\theta}{\partial \theta}(x)$ *and* $\frac{\partial^2 f_\theta}{\partial \theta^2}(x)$ *exists* $\forall x, \forall \theta \in \Theta$
- $\forall A$ *borelian set, the next integrals are well-defined and* $\frac{\partial}{\partial\theta}\int_A f_\theta(x)dx = \int_A \frac{\partial}{\partial\theta} f_\theta(x)dx$, $\frac{\partial^2}{\partial\theta^2}\int_A f_\theta(x)dx = \int_A \frac{\partial^2}{\partial\theta^2} f_\theta(x)dx$

Then the Fisher information is $I(\theta) = \mathbb{E}\left[\left(\frac{\partial \log f_\theta}{\partial\theta}(X)\right)^2\right]$

Efficient estimator

Theorem (Cramer-Rao bound)

Let θ_n unbiased estimator of θ, where Fisher information $I(\theta)$ is well-defined and non null, then

$$R(\theta_n, \theta) \geq \frac{1}{n} \frac{1}{I(\theta)}$$

Definition (Efficient estimator)

An estimator is called efficient if it reaches this lower bound.

Theorem (Efficiency of MLE)

Let θ_n MLE estimator of θ, under the same assumptions as the consistency theorem, then θ_n is efficient and

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(0, 1/I(\theta))$$

Some standard estimators

Definition (standard mean estimator)

Let $(x_1, \dots, x_n) \in \mathbb{R}^n$ sample supposedly generated by i.i.d. random variables of finite mean μ. The standard estimator of μ is the ***empirical mean****: $\overline{\mu}_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$. It is unbiased and strongly consistent.*

Definition (standard estimator of a finite discrete distribution)

Let $(x_1, \dots, x_n) \in E^n$ sample supposedly generated by i.i.d. random variables of discrete distribution over finite set E, with mass p_e for $e \in E$. The standard estimator of vector $p = (p_e)_{e \in E}$ is the ***frequency vector****: $\overline{p}_n(x_1, \dots, x_n) = \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{e\}}(x_i)\right)_{e \in E}$. It is an MLE, unbiased and strongly consistent.*

Some standard estimators

Definition (standard variance estimator when mean μ is known)

Let $(x_1,\ldots,x_n) \in \mathbb{R}^n$ sample supposedly generated by i.i.d. random variables of finite known mean μ and unknown variance V. The standard variance estimator of V is the ***empirical variance****: $\overline{V}_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$. It is unbiased and strongly consistent.*

Definition (unbiased standard variance estimator when mean μ is unknown)

Let $(x_1,\ldots,x_n) \in \mathbb{R}^n$ sample supposedly generated by i.i.d. random variables of finite unknown mean μ and unknown variance V. The standard variance estimator of V is the ***unbiased empirical variance****: $\overline{V}_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{\mu}_n)^2$ where $\overline{\mu}_n(x_1,\ldots,x_n) = \frac{1}{n-1}\sum_{i=1}^{n} x_i$.. It is unbiased and strongly consistent.*

Confidence Interval

Framework: statistical inference from a sample $(x_1, \ldots, x_n) \in E^n$ seen as a realization of a random vector $(X_1, \ldots, X_n)$ following a law with an unknown parameter $\theta \in \Theta$.

Definition (Confidence interval)

Let $0 < \alpha < 1$, consider two function I_n^- and I_n^+ from $\mathbb{R}^n$ to $\mathbb{R}$, if $\mathbb{P}(\theta \in [I_n^-(X_1, \ldots, X_n), I_n^+(X_1, \ldots, X_n)] = \alpha$ (resp. $\geq \alpha$), then this interval (whose extremities are random variables) is called a confidence interval for θ of exact level α (resp. of level α).

Extension: if this definition holds only when $n \to +\infty$, it is called an asymptotic confidence interval.

Confidence Interval

Example: suppose that the sample $(x_1, \ldots, x_n)$ has been generated by i.i.d. normal laws $\mathcal{N}(\mu, 1)$, find a confidence interval for μ of exact level α.

Consider the standard estimator of μ which is the empirical mean, then $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$, follows the normal law $\mathcal{N}(\mu, 1/n)$, thus $\sqrt{n}(\widehat{\mu}_n - \mu) \sim \mathcal{N}(0,1)$. For $\delta > 0$, we have:

$$\mathbb{P}(|\widehat{\mu}_n - \mu| \leq \frac{\delta}{\sqrt{n}}) = \frac{1}{\sqrt{2\pi}} \int_{-\delta}^{+\delta} e^{-x^2/2} dx$$

Given α, choose δ such that the integral equals α, then we can rewrite the inequalities and get:

$$\mathbb{P}(\mu \in [\widehat{\mu}_n - \delta/\sqrt{n}, \widehat{\mu}_n + \delta/\sqrt{n}]) = \alpha$$

Confidence Interval

Example: suppose that the sample $(x_1, \ldots, x_n)$ has been generated by i.i.d. normal laws $\mathcal{N}(\mu, \sigma^2)$ where σ is known and μ unknown, find a confidence interval for μ of exact level α.

Consider the standard estimator of μ which is the empirical mean, then $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, follows the normal law $\mathcal{N}(\mu, \sigma^2/n)$, thus $\sqrt{n}(\widehat{\mu}_n - \mu)/\sigma \sim \mathcal{N}(0,1)$. For $\delta > 0$, we have:

$$\mathbb{P}(|\widehat{\mu}_n - \mu| \leq \frac{\delta\sigma}{\sqrt{n}}) = \frac{1}{\sqrt{2\pi}} \int_{-\delta}^{+\delta} e^{-x^2/2} dx$$

Given α, choose δ such that the integral equals α, then we can rewrite the inequalities and get:

$$\mathbb{P}(\mu \in [\widehat{\mu}_n - \frac{\delta\sigma}{\sqrt{n}}, \widehat{\mu}_n + \frac{\delta\sigma}{\sqrt{n}}]) = \alpha$$

Confidence Interval

Example: suppose that the sample $(x_1, \dots, x_n)$ has been generated by i.i.d. normal laws $\mathcal{N}(\mu, \sigma^2)$ where σ is known and μ unknown, find a confidence interval for μ of exact level α.

Confidence interval for the mean of i.i.d. normal laws when the variance is known

1. choose the confidence level α
2. find the $(1+\alpha)/2$-quantile $q_{(1+\alpha)/2}$ of $\mathcal{N}(0,1)$
3. return the confidence interval $\mu \in [\widehat{\mu}_n - \frac{q_{(1+\alpha)/2}\sigma}{\sqrt{n}}, \widehat{\mu}_n + \frac{q_{(1+\alpha)/2}\sigma}{\sqrt{n}}]$ of exact level α, where $\mu_n(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Confidence Interval

Example: suppose that the sample $(x_1, \ldots, x_n)$ has been generated by i.i.d. normal laws $\mathcal{N}(\mu, \sigma^2)$ where σ is unknown and μ unknown, find a confidence interval for μ of exact level α.

Hint: consider the estimators for the mean $\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and for the variance $\overline{V}_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{\mu}_n)^2$, then $\frac{\overline{\mu}_n - \mu}{\sqrt{\overline{V}_n / n}} \sim t(n-1)$ the Student distribution with $n-1$ degrees of freedom.

Confidence interval for the mean of i.i.d. normal laws when the variance is unknown

1. choose the confidence level α
2. find the $(1+\alpha)/2$-quantile $q_{(1+\alpha)/2}$ of $t(n-1)$
3. return the confidence interval of exact level α, $\mu \in [\widehat{\mu}_n - \frac{q_{(1+\alpha)/2}\widehat{\sigma}_n}{\sqrt{n}}, \widehat{\mu}_n + \frac{q_{(1+\alpha)/2}\widehat{\sigma}_n}{\sqrt{n}}]$, where $\mu_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\sigma_n = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_n)^2\right)^{1/2}$

A decision problem

Framework: same as before but with a decision problem.

Example: let X random variable of uniform law in $[a, 1]$ where $0 \leq a < 1$ is unkown. A sample $(x_1, \ldots, x_n)$ has been generated by n independent trials of X. Can you find an algorithm which decides which is the right answer:

- H_0: $a = 0$
- H_1: $a > 0$

Ideas ?

Warning: two risks

- Reject H_0 whereas it is true (Type I error)
- Accept H_0 whereas it is false (Type II error)

A decision problem

Framework: same as before but with a decision problem.

Example: let X random variable of uniform law in $[a, 1]$ where $0 \leq a < 1$ is unkown. A sample $(x_1, \ldots, x_n)$ has been generated by n independent trials of X. Can you find an algorithm which decides which is the right answer:

- H_0: $a = 0$
- H_1: $a > 0$

Ideas ?

Warning: two risks

- Reject H_0 whereas it is true (Type I error) proba $\leq \beta$
- Accept H_0 whereas it is false (Type II error)

A decision problem

Framework: same as before but with a decision problem.

Example: let X random variable of uniform law in $[a,1]$ where $0 \leq a < 1$ is unkown. A sample $(x_1,\ldots,x_n)$ has been generated by n independent trials of X. Can you find an algorithm which decides which is the right answer:

- H_0: $a = 0$
- H_1: $a > 0$

Ideas ?

Warning: two risks

- Reject H_0 whereas it is true (Type I error) proba $\leq \beta$
- Accept H_0 whereas it is false (Type II error) try minimizing

A decision problem

Example: let X random variable of uniform law in $[a, 1]$ where $0 \leq a < 1$ is unkown. A sample $(x_1, \ldots, x_n)$ has been generated by n independent trials of X. Can you find an algorithm which decides which is the right answer: either H_0: $a = 0$, or H_1: $a > 0$

Idea: choose a threshold $s > 0$ and run the next algorithm

Test

1. if $\min(x_1, \ldots, x_n) < s$, accept H_0, else reject H_0

Question: how to choose s such that Type I error has proba $\leq \beta$?

A decision problem

Example: let X random variable of uniform law in $[a, 1]$ where $0 \leq a < 1$ is unkown. A sample $(x_1, \ldots, x_n)$ has been generated by n independent trials of X. Can you find an algorithm which decides which is the right answer: either H_0: $a = 0$, or H_1: $a > 0$
Idea: choose a threshold $s > 0$ and run the next algorithm

Test

1. if $\min(x_1, \ldots, x_n) < s$, accept H_0, else reject H_0

Question: how to choose s such that Type I error has proba $\leq \beta$?
Suppose $a = 0$, $\mathbb{P}(\text{Type I error}) = \mathbb{P}(\min(X_1, \ldots, X_n) \geq s) = (1-s)^n$. Thus choose s such that $(1-s)^n \leq \beta$, that is $1 - \beta^{1/n} \leq s \leq 1$. Now note that if $a > 0$, $\mathbb{P}(\text{Type II error}) = 0$ if $s \leq a$ and $\mathbb{P}(\text{Type II error}) = 1 - (\frac{1-s}{1-a})^n$ if $s > a$. To minimize this proba while ensuring low Type I error, choose $s = 1 - \beta^{1/n}$.

Chi-square test of goodness of fit

Hypothese: X random variable with values in $\{a(1), \dots, a(k)\}$

- H_0: X has vector $p = (p(1), \dots, p(k))$ as mass function
- H_1: X has another distribution

Question: given a sample $(x_1, \dots, x_n)$ generated by independant trials of X, provide an algorithm to decide H_0 with confidence level α (that is $\mathbb{P}(\text{Type I error}) \leq 1 - \alpha$).

Theorem

Let $f_n(i) \stackrel{def}{=} \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{a(i)\}}(x_j)$ *frequency of* $a(i)$ *in the sample.*
Let $\chi^2(p, f_n) \stackrel{def}{=} n \sum_{i=1}^{k} \frac{[p(i) - f_n(i)]^2}{p(i)}$.
Assuming H_0*, we have* $\chi^2(p, f_n) \xrightarrow{D} \chi^2(k-1)$ *(*χ^2*-distribution with* $k-1$ *degrees of freedom).*

Chi-square test of goodness of fit

Theorem

Let $f_n(i) \overset{def}{=} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{a(i)\}}(x_j)$ *frequency of* $a(i)$ *in the sample.*

Let $\chi^2(p, f_n) \overset{def}{=} n\sum_{i=1}^{k} \frac{[p(i)-f_n(i)]^2}{p(i)}$.

Assuming H_0*, we have* $\chi^2(p, f_n) \xrightarrow{D} \chi^2(k-1)$

Application: you throw a dice 120 times and you obtain the next output frequencies

Number	1	2	3	4	5	6
Frequency	14	16	28	30	18	14

Is this dice unbiased (hypothesis H_0) ? Answer with confidence level 95%

Chi-square test of goodness of fit

Theorem

Let $f_n(i) \stackrel{def}{=} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{a(i)\}}(x_j)$ *frequency of* $a(i)$ *in the sample.*

Let $\chi^2(p, f_n) \stackrel{def}{=} n\sum_{i=1}^{k} \frac{[p(i)-f_n(i)]^2}{p(i)}$.

Assuming H_0, *we have* $\chi^2(p, f_n) \xrightarrow{D} \chi^2(k-1)$

Application: you throw a dice 120 times and you obtain the next output frequencies

Number	1	2	3	4	5	6
Frequency	14	16	28	30	18	14

Is this dice unbiased (hypothesis H_0) ? Answer with confidence level 95%

$p = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, $\chi^2(p, f_n) \approx 12.8$ for sample, look at $\chi^2(5)$ table $\rightarrow$ 0.95-quantile $\approx 11.07 \rightarrow$ reject