Probability and Statistics Competitive Exam International Recruitment of IPParis Schools for Engineering Cycles

The notions of the mathematics test program (algebra, real analysis) are assumed to be known. The measure theory and abstract integration theory are out of the scope of this program (in particular all the measurability issues of functions).

1 Probability

1.1 Random Events

1. Sample space Ω, collection $\mathcal{E}$ of events : non empty collection of subsets of Ω closed under complement and countable unions.

2. Set operations on events : union, intersection, complement, difference...

3. Probability $\mathbb{P}$: function from $\mathcal{E}$ to $[0,1]$ such that $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $A \subset B$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if $A \cap B = \emptyset$ and $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$ if A_i is a sequence of disjoint sets ($A_i \cap A_j = 0$ for any $i \neq j$).

4. Conditional probability on a non negligible event B : $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$

5. $\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_1) \prod_{k=2}^n \mathbb{P}(A_k | A_1 \cap ... \cap A_{k-1})$

6. Law of total probability : $\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$ if $\mathbb{P}(\cup_i B_i) = 1$ and $\mathbb{P}(B_i \cap B_j) = 0$ for $i \neq j$

7. Bayes : $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$ and $\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$ if $\mathbb{P}(\cup_i A_i) = 1$ and $\mathbb{P}(A_i \cap A_j) = 0$ for $i \neq j$

8. Independence of collection of events and pair-wise independence of events.

9. Borel-Cantelli Lemma : if $(A_n)_{n\geq 0}$ is a sequence of events of $\mathcal{E}$ such that $\sum_{n\geq 0} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup_n A_n) = 0$.

1.2 Real random variables, random vectors

1. Definition of random variable : Function X from Ω in $\mathbb{R}$ such that $\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{E}$ for any a.
2. Notation $\mathbb{P}(X \in A)$, $\mathbb{P}(X > b)$, ... for $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$, $\mathbb{P}(\{\omega \in \Omega : X(\omega) > b\})$, ...
3. Cumulative distribution function : $F_X(a) = \mathbb{P}(X^{-1}(]-\infty; a])) = \mathbb{P}(X \leq a)$
4. Support $\mathcal{S}(X)$: smallest closed set F of $\mathbb{R}$ such that $\mathbb{P}(X \in F) = 1$.
5. Discrete random variable : random variable with a finite or countable support
6. Random variable admitting a density : random variable X with a cumulative distribution function admitting the integral representation $F_X(a) = \int_{-\infty}^{a} f_X(x)dx$, f_X is the probability density function of X
7. Random vector $X = (X_i)_{i=1,...,n}$, (multivariate) cumulative distribution function $F_X(x_1, ..., x_n) = \mathbb{P}(\cap_{i=1}^{n}\{X_i \leq x_i\})$, support $\mathcal{S}(X)$ (smallest closed set F of R^n such that $\mathbb{P}(X \in F) = 1$), discret case ($\mathcal{S}(X)$ finite or countable), multivariate density if F_X admits the integral representation $F_X(x_1, ..., x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f_X(x_1, ..., x_n)dx_1...dx_n$
8. Non negative random variable : random variable X such that $F_X(a) = 0$ for any $a < 0$
9. Expectation of $g(X)$ for g non negative function on $\mathcal{S}(X)$ and X random variable or vector : $\mathbb{E}(g(X)) = \sum_{x \in \mathcal{S}(X)} g(x)\mathbb{P}(X = x)$ if X is discrete, $\mathbb{E}(g(X)) = \int_{\mathcal{S}(X)} g(x)f_X(x)dx$ if X admits a density. For any functions g on $\mathcal{S}(X)$, the expectation of $g(X)$ is defined if $\mathbb{E}(|g(X)|) < \infty$ and in that case $E(g(X)) = \sum_{x \in \mathcal{S}(X)} g(x)\mathbb{P}(X = x)$ for X discrete, $\mathbb{E}(g(X)) = \int_{\mathcal{S}(X)} g(x)f_X(x)dx$ if X admits a density.
10. The expectation of a random vector is the vector of the expectation of each component.
11. Variance, covariance of random variables : $\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ and $\mathbb{C}ov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. Standard deviation of a random variable : $\sqrt{\mathbb{V}(X)}$.
12. Variance-covariance matrix of a random vector $(X_1, ..., X_n)$: square matrix of size $n \times n$ with components $\mathbb{C}ov(X_i, X_j)$
13. Markov Inequality : $\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$
14. Bienaymé-Tchebychev Inequality : $\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}$
15. Triangle Inequality: $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$
16. Indépendence of random variables $(X_i)_{i=1,...,n}$: we admit the equivalence of the following definition
    (a) $F_{(X_1,...,X_n)}(x_1, ..., x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$

(b) For any non negative functions g_i : $\mathbb{E}\left(\prod_{i=1}^n g_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}\left(g_i(X_i)\right)$

(c) For any bounded functions g_i : $\mathbb{E}\left(\prod_{i=1}^n g_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}\left(g_i(X_i)\right)$

(d) For any functions g_i such that $\mathbb{E}\left(\prod_{i=1}^n |g_i(X_i)|\right) < \infty$ and $\mathbb{E}\left(|g_i(X_i)|\right) < \infty$: $\mathbb{E}\left(\prod_{i=1}^n g_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}\left(g_i(X_i)\right)$

And when $(X_i)_{i,\dots,n}$ admits a density : $f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = \prod_{i=1}^n f_{X_i}(x_i)$.

17. The random variables $(X_i)_{i=1,\dots,n}$ are identically distributed if they share the same cumulative distribution.

18. The random variables $(X_i)_{i=1,\dots,n}$ are iid if they are independent and identically distributed

19. Sequence $(X_i)_{i\geq 1}$ of independent random variables (respectively identically distributed) : if for any $n \geq 2$, the variables $(X_i)_{1\leq i\leq n}$ are independent (respectively identically distributed)

20. Ability to compute moments of usual distributions : uniform, Gaussian, Poisson, Bernoulli, binomial, exponential, Gamma, ...

21. Simple example of transfer : from the density of X, applicants are expected be able to recover the density of $aX+b$, X^2, $\exp(X)$,...

1.3 Convergence of random variables

1. Almost-sure convergence of X_n to X: $\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$

2. Convergence in L^p (for $p \geq 1$): let a sequence o random variables X_n such that $\mathbb{E}(|X_n|^p) < \infty$ for any $n \geq 0$ and X a random variable such that $\mathbb{E}(|X|^p) < \infty$, X_n converges to X in L^p if $\mathbb{E}(|X_n - X|^p)$ tends to 0 when n tends to ∞.

3. Convergence in probability : the sequence of random variables X_n converges to the random variable X if for any $\epsilon > 0$ $\mathbb{P}(|X_n - X| > \varepsilon)$ tends to 0 when n tends to ∞.

4. Convergence in law or in distribution of X_n to X: At any continuity point x of F_X, $F_{X_n}(x)$ converges to $F_X(x)$. Other equivalent condition :$\mathbb{E}(f(X_n))$ converges to $\mathbb{E}(f(X))$ for any continuous and bounded f; or for any bounded Lipschitz f; or $\mathbb{E}(\exp(itX_n))$ converges to $\mathbb{E}(\exp(itX))$ for any $t \in \mathbb{R}$ (Levy continuity theorem).

5. Relation between the various convergences : convergence in L^p implies convergence in L^q for $p \geq q \geq 1$, convergence in L^p implies convergence in probability, almost-sure convergence implies convergence in probability, convergence in probability implies convergence in distribution. Theorem of dominated convergence to show the convergence in L^p when the almost-sure convergence holds.

6. Continuous mapping theorem for the convergence in probability, almost-sure, in distribution. If X_n converges to X in probability, (respectively in distribution, respectively almost-surely) and if g is a continuous function continue at any point of $\mathcal{S}(X)$ then $g(X_n)$ converges to $g(X)$ in probability (respectively in distribution, respectively almost-surely).

7. Slutsky Lemma: if X_n converges in distribution to X and Y_n converges in distribution to a constant c, then $X_n + Y_n$ converges in distribution to $X + c$ and $Y_n X_n$ converges in distribution to cX.

1.4 Statistics

1. Definition of a parametric model, parametrized by $\Theta \subset \mathbb{R}^K$: $\{(\mathbb{P}_\theta)_{\theta\in\Theta}, \Theta\}$, as a collection of probability distributions indexed by Θ.

2. Identification of $\theta \in \Theta$ in the statistical model: θ is identifiable if $\mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$ as soon as $\theta \neq \theta'$

3. Weak law of large numbers for variables with finite variance: if the X_i are iid variables, such that $\mathbb{E}(X_1^2) < \infty$ then $\frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to $\mathbb{E}(X_1)$. Be able to prove this with Bienaymé-Tchebychev inequality.

4. Strong law of large number for variables with finite expectation: if the X_i are iid variables, such that $\mathbb{E}(|X_1|) < \infty$ then $\frac{1}{n}\sum_{i=1}^n X_i$ converges almost-surely, in L^1 and in probability to $\mathbb{E}(X_1)$.

5. Central Limit Theorem : if X_i are iid variables, such that $\mathbb{E}(X_1^2) < \infty$ then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_1)\right) / \sqrt{\mathbb{V}(X_1)}$$

converges in distribution to the standard Gaussian distribution, in particular :

$$\mathbb{P}\left(a\sqrt{\mathbb{V}(X_1)} \leq \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_1)\right) \leq b\sqrt{\mathbb{V}(X_1)}\right)$$

converges to

$$\frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{t^2}{2}\right) dt.$$

6. Definition of an estimator : An estimator of $g(\theta) \in \mathbb{R}^k$ is a random variable (or a random vector if $k > 1$) of the form $T_n = \varphi_n(X_1, ..., X_n)$ where $X_1, ..., X_n$ are n random variables for which the realizations are observed. The function φ_n does not depend on θ

7. The bias of an estimator is $B = \mathbb{E}(T_n) - g(\theta)$, an unbiased estimator of $g(\theta)$ is an estimator T_n such that $\mathbb{E}(T_n) = g(\theta)$

8. Quadratic risk of an estimator T_n (for $g(\theta) \in \mathbb{R}$) : $R = \mathbb{E}\left(|T_n - g(\theta)|^2\right)$

9. Decomposition of the quadratic risk T_n (for $g(\theta) \in \mathbb{R}$) : $R = B^2 + V$, where V is the variance of T_n.

10. A consistent estimator is an estimator T_n that converges in probability to $g(\theta)$. A strongly consistent estimator is an estimator T_n that converges almost-surely to $g(\theta)$.

11. A confidence interval of $g(\theta) \in \mathbb{R}$ of confidence level $1-\alpha$ ($\alpha \in [0,1]$) of $g(\theta)$ is an interval $[U_n, V_n]$ such that :

   - U_n and V_n are some functions of observations $(X_1, ..., X_n)$ and of n but not of θ,
   - $\mathbb{P}(U_n \leq V_n) = 1$,
   - $\mathbb{P}(U_n \leq g(\theta) \leq V_n) \geq 1-\alpha$.

12. An asymptotic confidence interval of $g(\theta) \in \mathbb{R}$ of confidence level $1-\alpha$ ($\alpha \in [0,1]$) of $g(\theta)$ is a sequence of intervals $[U_n, V_n]$ such that :

   - U_n and V_n are some functions of observations $(X_1, ..., X_n)$ and of n but not of θ,
   - $\mathbb{P}(U_n \leq V_n) = 1$,
   - $\mathbb{P}(U_n \leq g(\theta) \leq V_n) \geq 1-\alpha_n$ for a sequence α_n tending to α.