Review on Probability Theory

Nima Reyhani

October 3, 2006

Let Ω be an abstract space, and $\mathcal{F} \subset 2^{\Omega}$. Suppose $\mathcal{F}$ satisfies

1. $\emptyset \in \mathcal{F}, \Omega \in \mathcal{F}$
2. $F \in \mathcal{F}$ implies that $F^c \in \mathcal{F}$
3. $\mathcal{F}$ is closed under finite intersection and union
4. $\mathcal{F}$ is closed under countable unions and intersections

$\mathcal{F}$ is an *algebra* if it satisfies (1),(2) and (3), and σ-*algebra* if it satisfies (1),(2) and (4).

σ algebra examples

- $\mathcal{A} = \{\emptyset, \Omega\}$
- $A \subset \Omega$, we have $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$
- when $\Omega = \mathbb{R}$, the *Borel σ-algebra* is the σ-algebra generated by open sets $(-\infty, a), a \in \mathbb{R}$.

Probability Measure $P : \mathcal{F} \rightarrow [0, 1]$ which satisfies

- $P(\Omega) = 1$
- **Coutable Addditivity** for coutable sequence A_n of elements $\mathcal{F}$ and $A_i \cap A_j = \emptyset$ we have

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

$P(A)$ is called the probability of A.

Random Variable

A random variable on probablity space $\{\Omega, \mathcal{F}, P\}$ is a function $x : \Omega \rightarrow X \subseteq \mathbb{R}$ such that $x^{-1}(B) \in \mathcal{F}, \forall B \in \mathcal{B}$. Probability of x is

$$P_x(B) = P\left(x^{-1}(B)\right), B \in \mathcal{B}$$

Distribution function

$$F_x(x) = P(\{\omega; x(\omega) \leq x\}) = P_x((-\infty, x)) \quad x \in \mathbb{R}$$

$$F_{\mathbf{x}}(\mathbf{x}) = F_{\mathbf{x}}(x_1, x_2, \ldots, x_k) = P_{\mathbf{x}}\{(-\infty, x_1) \cap \cdots \cap (-\infty, x_k)\}$$

Probability mass function

For discrete x define

$$p_x(x) = P(\{\omega; x(\omega) = x\})$$

Density function

$$P_x(x \in B) = \int_B p_x(t)dt = \int_B dF_x(t), \quad B \in \mathcal{B}$$

$$P_{\mathbf{x}}(B) = \int_B p_{\mathbf{x}} d\mathbf{x} = \int_B p_{\mathbf{x}}(x_1, x_2, \ldots, x_k)\, dx_1 dx_2 \ldots dx_k, B \in \mathcal{B}$$

Simple r.v. $X = \sum_1^n a_i 1_{A_i}$

Expectation $\mathbb{E}\left\{X\right\} = \sum_1^n a_i P\left(A_i\right) = \int X(\omega)P(d\omega) = \int X dp$

Let $X^+ = \max(0, X), X^- = \min(0, X)$, then r.v. X has a finite expectation if both $\mathbb{E}\left\{X^+\right\}, \mathbb{E}\left\{X^-\right\}$ and

$$\mathbb{E}\{X\} = \mathbb{E}\left\{X^+\right\} - \mathbb{E}\left\{X^-\right\}$$

Expectation of arbitrary r.v.

For every r.v. X there is a sequence $\{X_n\}_{n\leq 1}$ of possitive simple r.v. such that $X_n \rightarrow X$ as $n \rightarrow \infty$.

Example

$$X_n = \begin{cases} k2^{-n} & k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{and } 0 \leq k \leq n2^{-n} - 1 \\ n & X(\omega) \geq n \end{cases}$$

Moment of order $n = n_1 + n_2 + \cdots + n_k$

$$\mathbb{E}\left\{\prod_1^k x_i^{n_i}\right\}$$

Covariance

$$\mathbb{C}\left\{x_i, x_j\right\} = \mathbb{E}\{\left(x_i - \mathbb{E}\left\{x_j\right\}\right)\left(x_i - \mathbb{E}\left\{x_j\right\}\right)\}$$

Correlation

$$\mathbb{R}\left\{x_i, x_j\right\} = \frac{\mathbb{C}\left\{x_i, x_j\right\}}{\mathrm{Var}\left\{x_i\right\}^{\frac{1}{2}}\,\mathrm{Var}\left\{x_j\right\}^{\frac{1}{2}}}$$

Marginal Density

$(x_1, \ldots, x_k) \rightarrow \mathbf{y} = (x_1, \ldots, x_t) \quad \text{and} \quad \mathbf{z} = (x_{t+1}, \ldots, x_k)$

$$p_{\mathbf{y}} = \int_{\mathbb{R}^{k-t}} p(x_1, \ldots, x_k) dx_{t+1} \ldots dx_k$$

The conditional density $p_{\mathbf{z}|\mathbf{y}}(\mathbf{z}|\mathbf{y}) = \frac{p_{\mathbf{x}}(\mathbf{y},\mathbf{z})}{p_{\mathbf{y}}(\mathbf{y})}$

Bayes' Theorem

$$p_{\mathbf{y}|\mathbf{z}}(\mathbf{y}|\mathbf{z}) = \frac{p_{\mathbf{z}|\mathbf{y}}(\mathbf{z}|\mathbf{y}) p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{z}}(\mathbf{z})}$$

Independent random variables

- Sub σ-algebra $\{\mathcal{F}_i\}_{i\in I}$ of $\mathcal{F}$, are independent if for every finite subset $J \subseteq I$, and all $F_i \in \mathcal{F}_i$ one has

$$P\left(\cap_{i\in J A_i}\right) = \prod_{i\in J} P\left(A_i\right)$$

- R.V. $\{X_i\}$ with values in $(E_i, \mathcal{E}_i)$ are independnet if the generated σ-algebras $X_i^{-1}(\mathcal{E}_i)$ are independent

Conditional indepndence can be induced by restricting to the given information

Theorem

Two random variables X and Y are independent if and only if $\mathbb{E}\{f(X)g(Y)\} = \mathbb{E}\{f(X)\}\mathbb{E}\{g(Y)\}$ for every bounded $f$and g.

Characteristic function

$$\phi_x(t) = \mathbb{E}\left\{e^{itx}\right\}, t \in \mathbb{R}$$

Properties

- $|\phi_x(t)| = 1, \phi_x(0) = 1$
- $\phi_x(.)$ is uniformly continuous
- $x_1, \ldots, x_n$ iid, $s = \sum_1^n x_i$, then, $\phi_s(t) = \prod_1^n \phi_{x_i(}(t)$
- every distribution has it own characteristics function
- if $\mathbb{E}\left\{x^k\right\} < \infty$ then $\phi_x(t) = \sum_1^k \frac{(it))^j}{j!} \mathbb{E}\left\{x^j\right\} + o(t^k)$

Convergence type

In mean $\lim_{n\to\infty} \mathbb{E}\left\{(x_n - x)^2\right\} = 0$

Almost surely $P\left(\{\omega : \lim_{n\to\infty} x_n(\omega) = x(\omega)\}\right) = 1$

In probability $\forall\varepsilon > 0, \lim_{n\to\infty} P\left(\{\omega : |x_n(\omega) - x(\omega)| > \varepsilon)\right\} = 0$

In distribution $\lim_{n\to\infty} F_n(t) = F(t)$

Convergence in mean square	$\Longrightarrow$	Convergence in probability
Almost sure convergence	$\Longrightarrow$	Convergence in probability
Convergence in probability	$\Longrightarrow$	Convergence in ditribution

By definition, sequence f_t converges to f in limit, if the norm of difference converges to zero in limit

Suppose $x_1, x_2, \ldots, x_n$ are independent, identically distributed.

The "law of large numbers"

Let $\mathbb{E}\left\{x_i^2\right\} < \infty$ $\mathbb{E}\left\{x_i\right\} = \mu$ Then,

$$\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\}_n = \bar{x}_n \to \mu$$

in mean square sense

Weak law of large numbers Let $\mathbb{E}\left\{x_i\right\} = \mu$ Then, $\bar{x}_n$ converges in probability to μ

Strong law of large numbers Under the same condition as "weak law ...", $\bar{x}_n$ converges almost surely to μ

Suppose $x_1, x_2, \ldots, x_n$ are independent, identically distributed. The central limit theorem

central limit theorem when
$\mathbb{E}\{x_i\} = \mu \quad \mathbb{E}\left\{x_i^2\right\} - \mathbb{E}\{x_i\} = \sigma^2, \forall i$ then

$$z_n = \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$$

law of iterated logarithms Under the same condition as central limit theorem

$$\lim_{n\to\infty} \sup \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \left(2\log\log n\right)^{-\frac{1}{2}} = 1$$

Dirichlet Distribution

r.v. $\mathbf{x} = (x_1, \ldots, x_k)$ has a Dirchlet distribution, with parameters $\alpha = (\alpha_1, \ldots, \alpha_{k+1}) > 0$ if its probability density $Di(x|\alpha), 0 < x < 1, \sum_1^k < 1$ is

$$Di(x|\alpha) = \frac{\Gamma(\sum_1^{k+1} \alpha_i)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} x_1^{\alpha_1 - 1} \ldots x_k^{\alpha_k - 1} \left(1 - \sum_1^k x_i\right)^{\alpha_{k+1} - 1}$$

Then we also have

$$\mathbb{E}\{x_i\} = \frac{\alpha_i}{\sum_1^{k+1} \alpha_j}; \mathrm{Var}\{x_i\} = \frac{\mathbb{E}\{x_i\}(1 - \mathbb{E}\{x_i\})}{1 + \sum_1^{k+1} \alpha_j}$$

and

$$C\{x_i, x_j\} = -\frac{\mathbb{E}\{x_i\}\mathbb{E}\{x_j\}}{1 + \sum_1^{k+1} \alpha_j}$$

- $\alpha_i > 1, i = 1, \ldots, k$ there is mode given by

$$M\{x_i\} = \frac{\alpha_i - 1}{\sum 1^{k+1} \alpha_j - k - 1}$$

- Marginal distribution $x^{(m)} = (x_1, \ldots, x_m), m < k$ is

$$p(x^{(m)}) = Di_m\left(x^{(m)} | \alpha_1, \ldots, \alpha_m, \sum_{m+1}^{k+1} \alpha_j\right)$$

- The conditional distribution given $x_{m+1}, \ldots, x_k$ of

$$x_i' = \frac{x_i}{1 - \sum_{m+1}^{k} x_j}, i = 1, \ldots, m$$

is $\sim Di_m\left(x_1', x_2', \ldots, x_m' \mid \alpha_1, \ldots, \alpha_m, \alpha_{k+1}\right)$

Question

Find the exectation of

$$f(x) = \frac{1}{1+x^2}$$