1 Expected Value

1.1 Integration

- X is called a *simple function* on a measurable space $(\Omega, \mathcal{F})$ if $X = \sum_i a_i 1\{A_i\}$ is a finite sum, where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$.

- Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $a_i \geq 0$ for all i. The integral of X with respect to μ is

$$\int X \mathrm{d}\mu = \sum_i a_i \mu(A_i).$$

- Let X be a non-negative measurable function. The integral of X with respect to μ is

$$\int X \mathrm{d}\mu = \sup\left\{\int Y \mathrm{d}\mu : 0 \leq Y \leq X,\ Y \text{ is simple}\right\}.$$

- Let X be a measurable function. Define $X^+ = \max\{X, 0\}$ and $X^- = -\min\{X, 0\}$. Both X^+ and X^- are non-negative functions. The integral of X with respect to μ is

$$\int f \mathrm{d}\mu = \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu.$$

- If the measure μ is a probability measure P, then the integral $\int X \mathrm{d}P$ is called the *expected value,* or *expectation,* of X. We often use the popular notation $E[X]$, instead of $\int X \mathrm{d}P$, for convenience.

1.2 Properties

- Elementary calculation: $E[X] = \sum_x xP(X = x)$ or $E[X] = \int xf(x)\,\mathrm{d}x$.

- $E[1\{A\}] = P(A)$.

- $E[X^r]$ is call the r-moment of X. Mean $\mu = E[X]$, variance $\mathrm{var}[X] = E\left[(X-\mu)^2\right]$, skewness $E\left[(X-\mu)^3\right]$ and kurtosis $E\left[(X-\mu)^4\right]$.

- Skewness coefficient $E\left[(X-\mu)^3\right]/\sigma^3$, degree of excess $E\left[(X-\mu)^4\right]/\sigma^4 - 3$.
  - Application: The formula that killed Wall Street

- Jensen's inequality. If $\varphi(\cdot)$ is a convex function, then $\varphi(E[X]) \leq E[\varphi(X)]$.
  - Application: Kullback-Leibler distance $d(p, q) = \int \log(p/q)\,\mathrm{d}P = E_P[\log(p/q)]$

- Markov inequality: if $E\left[|X|^r\right]$ exists, then $P\left(|X|>\epsilon\right)\le E\left[|X|^r\right]/\epsilon^r$ for all $r\ge 1$.
    - Application: Chebyshev inequality: $P\left(|X|>\epsilon\right)\le E\left[X^2\right]/\epsilon^2$.

2 Multivariate Random Variable

- Bivariate random variable: $X:\Omega\mapsto\mathbb{R}^2$.
- Multivariate random variable $X:\Omega\mapsto\mathbb{R}^d$.
- Joint CDF: $F\left(x_1,\ldots,x_d\right)=P\left(X_1\le x_1,\ldots,X_d\le x_d\right)$. Joint PDF is defined similarly.

2.1 Law of Iterated Expectations

- Given a probability space $(\Omega,\mathcal{F},P)$, a sub σ-algebra $\mathcal{G}\subset\mathcal{F}$ and a $\mathcal{F}$-measurable function X with $E\left|X\right|<\infty$, the *conditional expectation* $E\left[X|\mathcal{G}\right]$ is defined as a $\mathcal{G}$-measurable function such that $\int_A XdP=\int_A E\left[X|\mathcal{G}\right]dP$ for all $A\in\mathcal{G}$.
- Law of iterated expectation

$$E\left[E\left[Y|X\right]\right]=E\left[Y\right]$$

  is a trivial fact from the definition of the conditional expectation by taking $A=\Omega$.
- Properties of conditional expectations
    1. $E\left[E\left[Y|X_1,X_2\right]|X_1\right]=E\left[Y|X_1\right]$
    2. $E\left[E\left[Y|X_1\right]|X_1,X_2\right]=E\left[Y|X_1\right]$
    3. $E\left[h\left(X\right)Y|X\right]=h\left(X\right)E\left[Y|X\right]$

2.2 Elementary Formulation

- conditional density $f\left(Y|X\right)=f\left(X,Y\right)/f\left(X\right)$
- marginal density $f\left(Y\right)=\int f\left(X,Y\right)dX$.
- conditional expectation $E\left[Y|X\right]=\int Yf\left(Y|X\right)dY$
- proof of law of iterated expectation
- conditional probability, or Bayes' Theorem

$$P\left(A|B\right)=\frac{P\left(A,B\right)}{P\left(B\right)}=\frac{P\left(B|A\right)P\left(A\right)}{P\left(B\right)}.$$

2.3 Application: Regression

- $Y = E[Y|X] + \epsilon$, where $\epsilon = Y - E[Y|X]$.

3 Independence

X and Y are *independent* if $P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$ for all A and B.

Application: (Chebyshev law of large numbers) If $X_1, X_2, \ldots, X_n$ are independent, and they have the same mean μ and variance $\sigma^2 < \infty$. Let $Z_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$. Then the probability $P(|Z_n| > \epsilon) \to 0$ as $n \to \infty$,