Useful Proofs for Probability and Statistics

April 2024

Useful Proofs

Theorem (Law of Total Probability). Suppose $\{B_1, B_2, \ldots\}$ is a partition of Ω by sets from $\mathcal{F}$, such that $\mathbb{P}(B_i) > 0$ for all $i \geq 1$. Then for any $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_{i \geq 1} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Proof. We have

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\left(A \cap \left(\bigcup_{i \geq 1} B_i\right)\right), \quad \text{since } \bigcup_{i \geq 1} B_i = \Omega \\
&= \mathbb{P}\left(\bigcup_{i \geq 1} (A \cap B_i)\right) \\
&= \sum_{i \geq 1} \mathbb{P}(A \cap B_i) \ \text{ by axiom 3, since } \forall i \neq j, (A \cap B_i) \cap (A \cap B_j) = \emptyset \\
&= \sum_{i \geq 1} \mathbb{P}(A|B_i)\mathbb{P}(B_i) \qquad \blacksquare
\end{aligned}$$

Theorem (Discrete Law of Iterated Expectations). For discrete values of X and Y,

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

Proof. We know that

$$\mathbb{E}[X] = \frac{1}{n} \sum_{i=0}^{n} x_i \mathbb{P}(X = x_i)$$

We must first prove the following lemma.
Lemma 1: If Y takes discrete values and thus any set of $\{Y_i\}$ is countable, then $\{Y_1, Y_2, \ldots\}$ is a partition of Ω.
Proof. By definition,

$$\Omega = \bigcup_{i \geq 1} Y_i$$

Given that for $i \neq j$, without loss of generality, assume that $Y_i = a$ and $Y_j = b$. We can then conclude that, since Y takes discrete values, that $a \neq b$. Hence, $Y_i \cap B_j = \emptyset$ $\blacksquare$

Proceeding with the main proof:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_y \mathbb{E}[X|Y=y]\mathbb{P}[Y=y] \\
&= \sum_y \sum_x x\mathbb{P}[X=x|Y=y]\mathbb{P}[Y=y] \\
&\implies \text{By Lemma 1 and the Law of Total Probability} \\
&= \sum_x x\mathbb{P}[X=x] \\
&= \mathbb{E}[X] \quad \blacksquare
\end{aligned}$$

Interesting Proofs

Theorem (Weak Law of Large Numbers). Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean μ. Then for any fixed $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - u\right| > \epsilon\right) \to 0$$

as $n \to 0$. Equivalently,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - u\right| \leq \epsilon\right) \to 1$$

To prove this, we need to prove two inequalities.
Theorem (Markov's inequality). Suppose that Y is a non-negative random variable whose expectation exists. Then,

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}, \forall t > 0$$

Proof. Let $A = \{Y \geq t\}$. We may assume that $\mathbb{P}(A) \in (0, 1)$, since otherwise the result is trivially true. Then by the law of total probability for expectations,

$$\mathbb{E}[Y] = \mathbb{E}[Y|A]\mathbb{P}(A) + \mathbb{E}[Y|A^c]\mathbb{P}(A^c) \geq \mathbb{E}[Y|A]\mathbb{P}(A)$$

since $\mathbb{P}(A^c) > 0$ and $\mathbb{E}[Y|A^c] \geq 0$. Now, we certainty have $\mathbb{E}[Y|A] = \mathbb{E}[Y|Y \geq t] \geq t$. So, rearranging, we get

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t} \quad \blacksquare$$

Theorem (Chebyshev's inequality). Suppose that Z is a random variable with a finite variance. Then for any $t > 0$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq \frac{Var(Z)}{t^2}$$

Proof. Note that $\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) = \mathbb{P}((Z - \mathbb{E}[Z])^2 \geq t^2)$, then apply Markov's inequality to the non-negative random variable $Y = Z - \mathbb{E}[Y]$.

Proof of Weak Law of Large Numbers (under the assumption of finite variance). Suppose the common distribution of the random variables X_i has mean μ and variance σ^2. Set

$$Z = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Since

$$\mathbb{E}[Z] = \mu \qquad \text{and} \qquad Var(Z) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}$$

So, by Chebyshev's inequality,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Since $\epsilon > 0$ is fixed, the right hand side tends to 0 as $n \to \infty$. $\blacksquare$