

Lecture 6

期望与方差-I

ruanyl@buaa.edu.cn

September 2016, Beihang University

6.1 随机变量的期望

离散随机变量及其函数的期望

离散随机变量可取至多可数个不同的实数, 因此它可以写成有限或可数求和

$$X(\omega) = \sum_i x_i I_{A_i}(\omega) \quad (6.1)$$

这里 $\bigcup_i A_i$ 是概率空间的划分. 例如若 x_i 是互不相同的, 可以将概率空间划分为 $A_i = \{\omega : X(\omega) = x_i\}$ 的并集. $I_A(\omega)$ 是集合 A 上的指示函数,

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in A^c \end{cases}.$$

如果 X 仅取有限个值, 那么我们称 X 是**简单随机变量**.

定义 6.1.1 假设 X 是 (6.1) 中的离散随机变量. 定义如下“加权

平均值”

$$\sum_i x_i \cdot P(A_i) = x_1 P(A_1) + x_2 P(A_2) + x_3 P(A_3) + \cdots$$

如果这个求和是绝对收敛的, 那么我们把这个求和称为 X 的期望, 并且记作

$$E(X) = \sum_i x_i \cdot P(A_i) = \sum_i x_i \cdot P(X = x_i)$$

注意 $E(X)$ 的定义仅仅依赖于 X 的分布, 因此如果 X, Y 具有相同分布, 那么 $E(X) = E(Y)$.

定理 6.1.1 假设 $g(x)$ 是实数空间上某函数, X 是 (6.1) 中的离散随机变量. 那么 $Y = g(X)$ 的期望是 (如果该求和绝对收敛)

$$E(Y) = \sum_i g(x_i) \cdot P(X = x_i)$$

■ 事实上

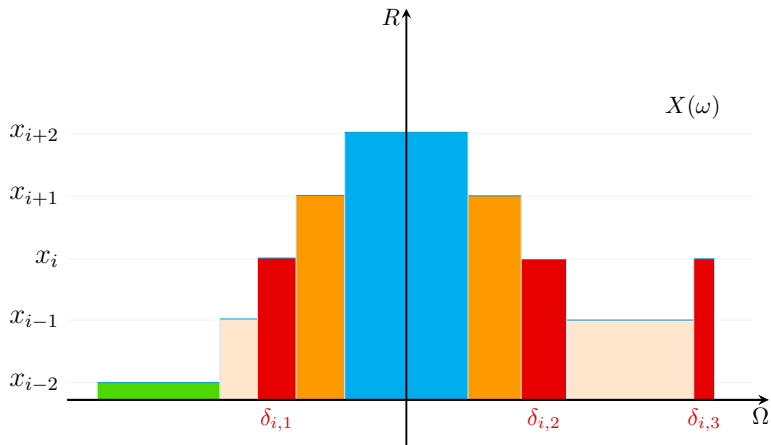
$$Y(\omega) = \sum_x g(x_i) \cdot I_{A_i}(\omega),$$

而 $\bigcup_i A_i = \bigcup_i \{X = x_i\}$ 形成概率空间的划分, 因此由定义可知 Y 的期望是

$$\sum_i g(x_i) \cdot P(A_i) = \sum_i g(x_i) \cdot P(X = x_i).$$



离散随机变量在概率空间的积分可以通过下图来理解



$$P(\omega : X(\omega) = x_i) = P(\delta_{i,1}) + P(\delta_{i,2}) + P(\delta_{i,3})$$

Figure 6.1: 简单随机变量积分示意图

连续型随机变量及其函数的期望

我们已经知道如何求简单随机变量的积分. 为了求连续型随机变量的积分, 我们将用简单随机变量来近似连续型随机变量, 进而用简单随机变量的积分来近似连续型随机变量的积分.

给定随机变量 X , 将其取值空间 \mathbb{R} 分割为互不相交的小区间 $B_i = [x_i, x_{i+1})$, $\bigcup_i B_i = \mathbb{R}$. 小区间的度量为

$$|B_i| = \Delta x_i = x_{i+1} - x_i.$$

如果划分充分细小, 我们可以认为随机变量 X 在 B_i 上近似地取常数值 $x_i \in B_i$. 例如, 如果 $X \geq 0$, 可取 $B_i = [\frac{i}{2^n}, \frac{i+1}{2^n})$, 令

$$X_n(\omega) = \begin{cases} \frac{i}{2^n}, & X(\omega) \in [\frac{i}{2^n}, \frac{i+1}{2^n}), 0 \leq i \leq n2^n - 1 \\ n, & X(\omega) \geq n \end{cases}$$

那么 X_n 是一列收敛 ($n \rightarrow \infty$) 到 X 的简单随机变量.

注意, 如果 $\bigcup_i B_i$ 是 \mathbb{R} 的一个划分, 令 $\Omega_i = \{\omega : X(\omega) \in B_i\}$, 那么 $\bigcup_i \Omega_i$ 是 Ω 的一个划分.

一般随机变量在概率空间的积分可以通过下图来理解

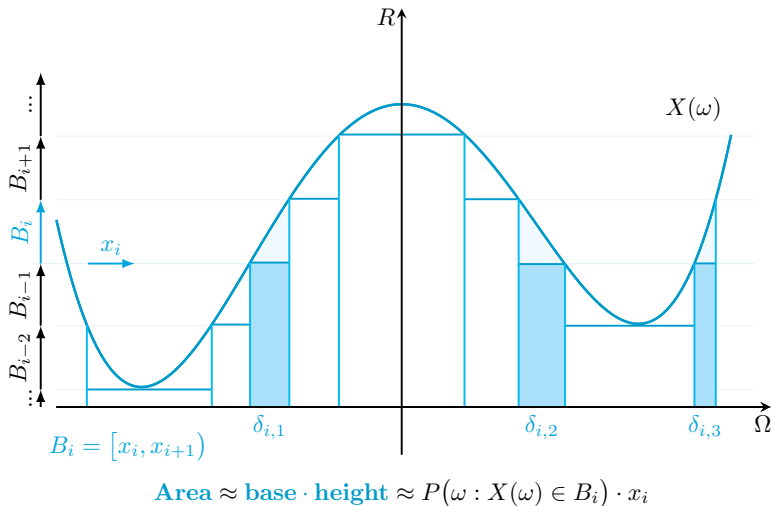


Figure 6.2: 随机变量积分示意图

因此我们只需要对简单随机变量 X_n 计算积分, 而 X_n 的积分 (参见6.2):

$$E(X_n) = \sum_i P(X_n = x_i) \cdot x_i = \sum_i P(X \in B_i) \cdot x_i$$

如果 X 的密度函数为 $f(x)$, 那么

$$\begin{aligned} E(X_n) &= \sum_i P(X \in B_i) \cdot x_i \\ &= \sum_i \int_{-\infty}^{\infty} I_{B_i}(x) f(x) dx \cdot x_i \\ &= \int_{-\infty}^{\infty} \left(\sum_i x_i I_{B_i}(x) \right) f(x) dx \\ &= \int_{-\infty}^{\infty} \left(\sum_i x_i I_{B_i}(x) \right) f(x) dx \end{aligned}$$

当 $n \rightarrow \infty$, 划分 $\bigcup_i B_i$ 变得越来越细,

$$\sum_i x_i I_{B_i}(x) \rightarrow x$$

并且

$$E(X_n) \rightarrow \int_{-\infty}^{\infty} xf(x) dx$$

因而我们有

定义 6.1.2 假设随机变量 X 的密度函数为 $f(x)$, 如果存在 (绝对收敛) 积分

$$\int_{-\infty}^{\infty} xf(x) dx$$

那么把它称为 X 的期望, 并记作

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

定理 6.1.2 假设 $g(x)$ 是实数空间上某函数, X 是以 $f(x)$ 为密度函数的连续型随机变量. 那么 $Y = g(X)$ 的期望是 (如果该积分绝对收敛)

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x) dx$$

■ 如同前面, 将 $g(x)$ 取值空间 \mathbb{R} 分割为互不相交的小区间 $B_i = [y_i, y_{i+1})$, $\bigcup_i B_i = \mathbb{R}$. 小区间的度量为

$$|B_i| = \Delta y_i = y_{i+1} - y_i.$$

令 $C_i = g^{-1}(B_i)$. 于是 Y 可以由 $\bigcup_i C_i$ 上的简单随机变量 Y_n 近似 (参见6.3) 而 $E(Y)$ 可以用 $E(Y_n)$ 来近似

$$E(Y_n) = \sum_i P(Y_n = y_i) \cdot y_i$$

$$\begin{aligned}
&= \sum_i P(Y \in B_i) \cdot y_i \\
&= \sum_i P(X \in g^{-1}(B_i)) \cdot y_i \\
&= \sum_i \int_{-\infty}^{\infty} I_{C_i}(x) f(x) dx \cdot y_i \\
&= \int_{-\infty}^{\infty} \left(\sum_i y_i I_{C_i}(x) \right) f(x) dx
\end{aligned}$$

根据 Y_n 的定义,

$$\sum_i y_i I_{C_i}(x) \rightarrow g(x)$$

因而上面的求和收敛到

$$E(Y_n) \rightarrow \int_{-\infty}^{\infty} g(x) f(x) dx$$



对多个随机变量, 同样的结论成立, 例如已知随机向量 $\mathbf{Z} = (X, Y)$ 有分布密度 $f_{\mathbf{Z}}(\mathbf{z}) = f(x, y)$, $g(\mathbf{z}) = g(x, y)$ 是定义在 \mathbb{R}^2 上的函数, 那么随机变量 $g(\mathbf{z})$ 的期望 (若存在) 为

$$E(\mathbf{Z}) = \begin{cases} \int_{\mathbb{R}^2} g(\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}, & \text{连续型} \\ \sum_{\mathbf{z}} g(\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}), & \text{离散型} \end{cases}$$
$$= \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy, & \text{连续型} \\ \sum_x \sum_y g(x, y) f(x, y), & \text{离散型} \end{cases}$$

随机变量的函数在概率空间的积分可以通过下图来理解

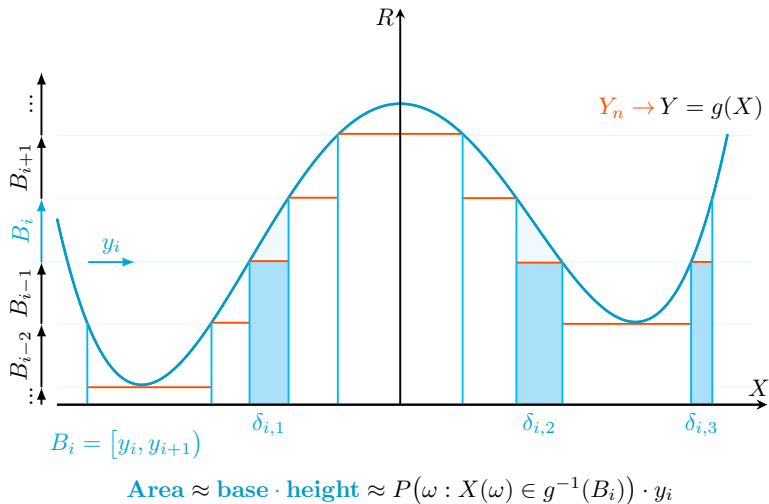


Figure 6.3: 随机变量函数的积分示意图

例题 6.1.1 已知 X 的密度为

$$f(x) = \begin{cases} \frac{3}{8}x^2, & 0 < x < 2 \\ 0, & \text{其他} \end{cases}$$

试求 $\frac{1}{X^2}$ 的期望.

■ $\frac{1}{X^2}$ 的期望为

$$E\left(\frac{1}{X^2}\right) = \int_{-\infty}^{\infty} \frac{1}{x^2} f(x) dx = \int_0^2 \frac{1}{x^2} \cdot \frac{3}{8} x^2 dx = \frac{3}{4}$$



条件期望

在条件概率一节我们看到给定 $X = x$, $f_Y(\cdot|x)$ 是一个概率密度, $g(x)$ 是实数空间上某函数, 因此可以考虑随机变量 $g(Y)$ 在

给定 $X = x$ 时的期望, 称之为**条件期望**, 写作

$$E(g(Y) | X = x) = \begin{cases} \int_{-\infty}^{\infty} g(y) f_Y(y|x) dy, & \text{连续型} \\ \sum_y g(y) f_Y(y|x), & \text{离散型} \end{cases}$$

此时它是一个 x 的函数, 对上式同乘以 $f_X(x)$ 并积分 (求和) 便得到

$$E(g(Y)) = \begin{cases} \int_{-\infty}^{\infty} E(g(Y) | X = x) f_X(x) dx, & \text{连续型} \\ \sum_x E(g(Y) | X = x) f_X(x), & \text{离散型} \end{cases} \quad (6.2)$$

这是条件期望的性质, 常常用于期望的计算.

6.2 期望与方差及其性质

假设 X 为随机变量, g 是 X 的值域空间上某函数, 我们已经知道如何求 $Y = g(X)$ 的期望. 下面我们考虑几个 $g(x)$ 的情形.

(1) $n \geq 1$ 为正整数, $g(x) = x^n$, n 阶矩 $E(X^n)$, 1 阶矩正是期望

(2) $g(x) = (x - E(X))^2$, 方差 $D(X) = E[(x - E(X))^2]$.

期望及其性质

我们已经知道, 随机变量 X 的期望就是 X 在概率空间上的积分, 它具有如下性质:

定理 6.2.1 假设 X, Y 为随机变量, 以下成立

(1) $E(I_A) = P(A)$, $A \subset \mathbb{R}$;

(2) $E(\alpha) = \alpha$, $\alpha \in \mathbb{R}$.

(3) 如果 EX, EY 存在, 那么 $E(\alpha X + \beta Y)$ 存在, $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$, $\alpha, \beta \in \mathbb{R}$.

(4) $X(\omega) \leq Y(\omega), \forall \omega$. 那么 $E(X) \leq E(Y)$.

(5) $|E(X)| \leq E|X|$.

注记 6.2.1 如果除去一个概率为 0 的集合外, $X \leq Y$ 成立, 那么结论 (4) 仍然成立.

■ 证明简单随机变量情形. (1)(2) 为期望的特殊情形, 可直接由定义得出.

(3) 假设

$$X = \sum_i x_i I_{A_i}, Y = \sum_j y_j I_{B_j}$$

首先注意 $\cup A_i, \cup B_j$ 都是全空间的划分, 因此

$$E(X) = \sum_i x_i P(A_i) = \sum_{i,j} x_i P(A_i \cap B_j),$$

$$E(Y) = \sum_j y_j P(B_j) = \sum_{i,j} y_j P(A_i \cap B_j).$$

另外 $\alpha X + \beta Y = \sum_{i,j} (\alpha x_i + \beta y_j) I_{A_i \cap B_j}$, 因此

$$\begin{aligned} E(\alpha X + \beta Y) &= \sum_{i,j} (\alpha x_i + \beta y_j) P(A_i \cap B_j) \\ &= \alpha \sum_{i,j} x_i P(A_i \cap B_j) + \beta \sum_{i,j} y_j P(A_i \cap B_j) \\ &= \alpha \sum_i x_i P(A_i) + \beta \sum_j y_j P(B_j) \\ &= \alpha E(X) + \beta E(Y) \end{aligned}$$

(4) 假设 X, Y 可如同 (3) 表示出来. 由假设当 $A_i \cap B_j$ 不空时, 总有 $x_i \leq y_j$. 因此 $E(X) \leq E(Y)$.

(5) 由于 $-|X| \leq X \leq |X|, \forall \omega$. 运用 (1) 便有

$$-E(|X|) \leq E(X) \leq E(|X|)$$

即 $|E(X)| \leq E|X|$.



通过将随机变量分解为简单随机变量的和, 期望的可加性质常常用于计算较复杂随机变量的期望.

例题 6.2.1 对目标进行射击, 直到击中目标 n 次为止, 如果每次射击命中率为 p ($0 < p < 1$), 求所需射击次数的期望.

■ (解法 I) 用 X_i 表示击中目标总数为 i 次时所需要的射击次数. 那么 $X_i - X_{i-1}$ 表示第 $i-1$ 次击中目标后再次击中目标所用的射击次数. 容易看出 $Y_i = X_i - X_{i-1}$, $i \geq 1$, ($X_0 = 0$) 相互独立且与 Y_1 有同样的分布:

$$P(Y_1 = k) = p(1-p)^{k-1}, k = 1, 2, \dots$$

因此

$$EY_1 = \sum_{k \geq 1} kP(Y_1 = k) = p \sum_{k \geq 1} k(1-p)^{k-1}$$

$$\begin{aligned}
 &= p \left(\sum_{k \geq 0} x^k \right)' \bigg|_{x=1-p} = p \left(\frac{1}{1-x} \right)' \bigg|_{x=1-p} \\
 &= p \cdot \frac{1}{p^2} = \frac{1}{p}
 \end{aligned}$$

从而

$$EX_n = E \left(\sum_{i=1}^n Y_i \right) = \sum_{i=1}^n EY_i = \frac{n}{p}$$

(解法 II) 用 X_n 表示击中目标总数为 n 次时所需要的射击次数. 令 $W_1 = 1$ 表示第一次射击击中目标, $W_1 = 0$ 表示第一次射击没有击中目标. 由 (6.2)

$$EX_n = E(X_n | W_1 = 1) P(W_1 = 1) + E(X_n | W_1 = 0) P(W_1 = 0)$$

首先注意各次射击是独立的. 如果第一次击中目标, 那么还需要再击中目标 $n - 1$ 次才能实现击中目标总次数为 n 次. 如果


第一次没有击中目标, 那么第一次射击对总射击次数 n 次没有贡献, 从第二次射击开始还需要再击中目标 n 次. 因此上面的方程变为:

$$EX_n = p(1 + EX_{n-1}) + (1 - p)(1 + EX_n)$$

边界条件为 $EX_0 = 0$, 解之得到 $EX_n = EX_{n-1} + 1/p$, 从而 $EX_n = n/p$.

(解法 III) 用 X_n 表示击中目标总数为 n 次时所需要的射击次数. 令 $Z_1 = 1$ 表示第 $n - 1$ 次击中目标后, 下一次射击击中目标, $Z_1 = 0$ 表示第 $n - 1$ 次击中目标后, 下一次射击没有击中目标. 注意由独立性可知 $EZ_1 = EX_1$. 类似上一解法我们有

$$EX_n = p(EX_{n-1} + 1) + (1 - p)(EX_{n-1} + 1 + EZ_1)$$

边界条件为 $EX_0 = 0$, 先求得 $EZ_1 = EX_1 = 1/p$, $EX_n = EX_{n-1} + 1/p$, 解得 $EX_n = n/p$. 

定理 6.2.2 假设 X 为**非负**随机变量, 那么

$$E(X) = \int_0^{\infty} P(X \geq t) dt$$

■ 仅证明非负简单随机变量情形. 令 X 可取值为 $0 \leq x_1 < x_2 < \dots < x_n$, 那么

$$\begin{aligned} E(X) &= \sum_{x_i} x_i \cdot P(X = x_i) \\ &= \sum_{x_i} \int_0^{x_i} dt \cdot P(X = x_i) \\ &= \int_0^{\infty} \sum_{x_i \geq t} P(X = x_i) dt \quad (\text{交换积分与求和次序}) \end{aligned}$$

$$= \int_0^{\infty} P(X \geq t) dt$$

由于 X 为简单随机变量, 这里的无穷区间积分 $\int_0^{\infty} (\cdot)$ 实际上是有限区间积分. //

注记 6.2.2 事实上仅有至多可数个 t 使得 $P(X > t) \neq P(X \geq t)$, 因而该定理可进一步写作: X 为**非负**随机变量, 那么

$$E(X) = \int_0^{\infty} P(X \geq t) dt = \int_0^{\infty} P(X > t) dt.$$

如果 X 的分布函数为 F , 那么

$$E(X) = \int_0^{\infty} (1 - F(t)) dt.$$

另外如果 X 取值为 $\{0, 1, 2, \dots\}$, 那么结论进一步简化为

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n).$$

运用该结论我们再来看之前的例题

例题 6.2.2 对目标进行射击, 直到击中目标为止, 如果每次射击命中率为 p ($0 < p < 1$), 求所需射击次数的期望.

■ 用 X 表示需射击次数. 注意 $\{X \geq n\}$ 等价于前 $n-1$ 次没有击中目标, 其概率为

$$P(X \geq n) = (1-p)^{n-1}$$

因此

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{p}.$$



定理 6.2.3 假设 X_1, X_2, \dots, X_n 为独立随机变量, $E|X_i| < \infty, \forall i$, 那

么乘积 $X_1 X_2 \cdots X_n$ 的期望存在, 并且

$$E(X_1 X_2 \cdots X_n) = E(X_1) E(X_2) \cdots E(X_n)$$

■ 证明 $n = 2$ 时简单随机变量的情形. 假设

$$X = \sum_i x_i I_{A_i}, Y = \sum_j y_j I_{B_j}$$

那么由独立性

$$E(XY) = \sum_{i,j} x_i y_j P(A_i \cap B_j) = \sum_{i,j} x_i y_j P(A_i) P(B_j) = E(X) E(Y)$$



几个不等式

// (Jensen 不等式) ϕ 为凸函数, $E[X], E[\phi(X)]$ 存在, 那么

$$\phi(E[X]) \leq E[\phi(X)]$$

// (Holder 不等式) X, Y 为随机变量, $p > 1, q > 1, \frac{1}{p} + \frac{1}{q} = 1$. 那么

$$E(|XY|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}$$

$p = q = 2$ 时称为 Schwarz 不等式. 特别地, 如果 $E(|X|^p) < \infty, E(|Y|^q) < \infty$ 那么 $E(|XY|) < \infty$. 此时, Holder 不等式等号成立当且仅当存在 $(a, b) \neq (0, 0)$, 使得 $P(a|X|^p + b|Y|^q = 0) = 1$, 即 $a|X|^p + b|Y|^q$ 几乎处处为 0.

// (Minkowski 不等式) X, Y 为随机变量, $p \geq 1$. 那么

$$(E(|X + Y|^p))^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}$$

// (Lyapounov 不等式) $0 < \alpha \leq \beta$. 那么

$$(E(|X|^\alpha))^{1/\alpha} \leq \left(E(|X|^\beta)\right)^{1/\beta}$$

// (Markov 不等式) $0 < \alpha$, 那么

$$P(|X| \geq \alpha) \leq \frac{1}{\alpha} E(|X|)$$

// (Chebyshev 不等式) $0 < \alpha$, $E(X^2) < \infty$. 令 $m = E(X)$, 那么

$$P(|X - m| \geq \alpha) \leq \frac{1}{\alpha^2} E(|X - m|^2)$$

这里多数不等式都与函数积分不等式平行, 不再重复证明. Schwarz 不等式及其等号成立条件将会在协方差一节给出证

明. Chebyshev 不等式与 Markov 不等式的证明相同, 这里仅证明 Markov 不等式. 事实上, 令 $A = \{|X| \geq \alpha\}$, 那么

$$|X(\omega)| = |X(\omega)| \cdot 1_A(\omega) + |X(\omega)| \cdot 1_{A^c}(\omega), \forall \omega$$

于是

$$\begin{aligned} E(|X|) &= E(|X| \cdot 1_A) + E(|X| \cdot 1_{A^c}) \\ &\geq E(|X| \cdot 1_A) \\ &\geq E(\alpha \cdot 1_A) = \alpha P(|X| \geq \alpha) \end{aligned}$$

因此 Markov 不等式成立.

作为 Markov 不等式的简单运用, 我们有

定理 6.2.4 假设 X 为非负随机变量, $E(X) = 0$ 当且仅当 $P(X = 0) = 1$.

■ (必要性) 令 $A_n = \{X \geq \frac{1}{n}\}$, 那么由 Markov 不等式, $\forall n$,

$$P\left(X \geq \frac{1}{n}\right) \leq nE(X) = 0$$

即 $P(A_n) = 0, \forall n$. 注意到 $\cup_n [\frac{1}{n}, \infty) = (0, \infty)$, $\cup_n A_n = \{X > 0\}$.
由概率的下半连续性就得到

$$P(X > 0) = \lim_n P(A_n) = 0$$

从而 $P(X = 0) = 1$.

(充分性) 证明简单非负随机变量情形. 设 $X = \sum_i x_i I_{A_i}$, $x_i \geq 0$. 注意 $P(X = 0) = 1$ 等价于, 对任意 i , 都有 $x_i = 0$ 或者 $P(A_i) = 0$. 因此 $E(X) = 0$. 