分布の再生性

Review

Th

X, Y : r.v., independent

$\Rightarrow {}^{\forall} h_X(X),\ h_Y(Y),$

$$E[h_X(X) \cdot h_Y(Y)] = E[h_X(X)]E[h_Y(Y)]$$

Cor

X, Y : r.v., independent

$\Rightarrow$ ① $E[XY] = E[X]E[Y]$

② $M_{X+Y}(t) = M_X(t) M_Y(t)$

②の証明

$$M_{X+Y}(t) = E[e^{t(X+Y)}]$$
$$= E[e^{tX} \cdot e^{tY}]$$
$$= E[e^{tX}] \cdot E[e^{tY}] = M_X(t) \cdot M_Y(t).$$

//

Th

X : r.v.

$a \in \mathbb{R}$

$\Rightarrow M_{(aX)}(t) = M_X(at)$

Proof

$$M_{(aX)}(t) = E[e^{t(aX)}]$$
$$= E[e^{(at)X}]$$
$$= M_X(at).$$

//

Th

$$X_1, \cdots, X_n \sim iid(\mu, \sigma^2)$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Rightarrow E[\bar{X}_n] = \mu$$

$$V[\bar{X}_n] = \frac{\sigma^2}{n}$$

標本平均
(sample mean)

Th

X : r.v.

$$E[X] = \mu$$

$$V[X] = \sigma^2 > 0$$

$$Z_X = \frac{X-\mu}{\sigma}$$

標準化確率変数

$$\Rightarrow E[Z_X] = 0$$

$$V[Z_X] = 1$$

$$E[Z_X^2] = 1$$

分布の再生性

$X+Y$ の分布について

モーメント母関数 $M(t)=E[e^{tX}]$ を利用する。

Th

X, Y : r.v.

$M_X(t) = M_Y(t)$

$\Longrightarrow F_X = F_Y$

where F_X : distribution f. for X
F_Y : 〃 Y

Proof

We omit it here. //

※モーメント母関数が存在することは前提。

※モーメント母関数が一致するならば、同一の分布である。

系 (二項分布の再生性)

$$
\boxed{
\begin{array}{l}
X \sim B(m, p) \\
Y \sim B(n, p) \\
X, Y : \text{independent} \\
\Rightarrow X + Y \sim B(m+n, p)
\end{array}
}
$$

Proof

We show that $\underline{M_{X+Y}(t) = (pe^t + 1 - p)^{m+n}}$.

It holds that

$$
\begin{aligned}
& M_{X+Y}(t) \\
& \quad = M_X(t) \cdot M_Y(t) \qquad \Big\} \; X, Y : \text{independent} \\
& \quad = (pe^t + 1 - p)^m \cdot (pe^t + 1 - p)^n \\
& \quad = (pe^t + 1 - p)^{m+n}.
\end{aligned}
$$

$\therefore X + Y \sim B(m+n, p)$.

//

$$
\boxed{
\begin{array}{l}
X \sim B(n, p) \\
\Leftrightarrow M_X(t) = (pe^t + 1 - p)^n
\end{array}
}
$$

例

一年以内の事故確率が $\frac{1}{100}$ の契約

昨日 150件
今日 50件
の新規契約

(1) 昨日の契約の中から、一年以内に事故が起こる確率を求めよ。

X：昨日の契約の中からの一年以内の事故件数, r.v.

$X \sim B\left(150, \frac{1}{100}\right)$ と考えられるので

X の確率関数 P_X は

$$P_X(x) = {}_{150}C_x \left(\frac{1}{100}\right)^x \left(\frac{99}{100}\right)^{150-x}$$

$$(x = 0, 1, \cdots, 150)$$

である。

$$P_X(0) = {}_{150}C_0 \left(\frac{1}{100}\right)^0 \left(\frac{99}{100}\right)^{150} \fallingdotseq 22.145\%$$

よって事故が起こる確率は、

$$1 - P_X(0) = \underline{約77.85\%}$$

(2) 今日契約した中から一年以内に事故が起こる確率は？

Y：今日契約した中からの一年以内の事故契約数，r.v.

とする。

先ほどと同様に

$Y \sim B\left(50, \frac{1}{100}\right)$ なので

$$P_Y(y) = {}_{50}C_y \left(\frac{1}{100}\right)^y \left(\frac{99}{100}\right)^{50-y}$$

$$(y = 0, 1, 2, \cdots, 50)$$

である。よって

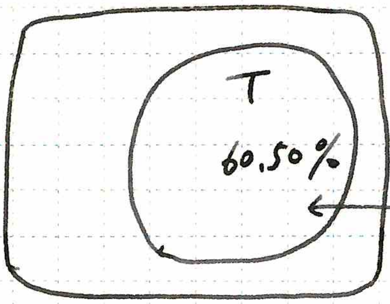
$$P_Y(0) = {}_{50}C_0 \left(\frac{1}{100}\right)^0 \left(\frac{99}{100}\right)^{50} \fallingdotseq 60.50\%$$

これより，事故が起こる確率は

$$1 - P_Y(0) \fallingdotseq 39.50\%$$

約39.50%。

(1)と比べて
契約件数が少ない
だけに事故が起こる
確率は低くなる。
しかし，1/3になるわけではない。



今日の契約から
事故ゼロ

(3) この2日間の契約から
事故が起こる確率は?

$Z = X + Y$ とする。
これがこの2日間の契約からの事故件数
を表す r.v. である。
二項分布の再生性により

$$Z \sim B\left(200, \frac{1}{100}\right)$$

である。よって

$$P_Z(z) = {}_{200}C_z \left(\frac{1}{100}\right)^z \left(\frac{99}{100}\right)^{200-z}$$

$(z = 0, 1, 2, \cdots, 200)$

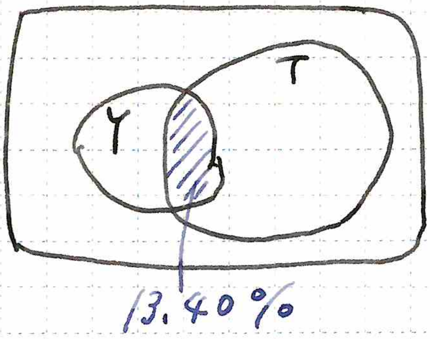
となる。これより

$$P_Z(0) = {}_{200}C_0 \left(\frac{1}{100}\right)^0 \left(\frac{99}{100}\right)^{200} \fallingdotseq 0.134。$$

約13.40%

よって事故が起こる確率は <u>約86.60%</u>.

77.85% + 39.50% (>100%)
(1)の結果　(2)の結果
では<u>ない</u>。



Thm (ポワソン分布の再生性)

$X \sim Po(\lambda_1)$
$Y \sim Po(\lambda_2)$
X, Y: independent
$\Rightarrow X+Y \sim Po(\lambda_1+\lambda_2)$

Proof

We prove that $M_{X+Y}(t) = e^{(\lambda_1+\lambda_2)(e^t-1)}$.

It holds that

$$
\begin{aligned}
&M_{X+Y}(t) \\
&= M_X(t) M_Y(t) \\
&= e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} \\
&= e^{(\lambda_1+\lambda_2)(e^t-1)} .
\end{aligned}
$$

//

$X \sim Po(\lambda)$
$\Leftrightarrow M_X(t) = e^{\lambda(e^t-1)}$

例. あるデパートの売り場

X_1 : 一時間あたりの来客数, r.v.

$X_1 \sim Po(2)$

確率関数 $p_1(x) = e^{-2} \dfrac{2^x}{x!}$

$(x=0,1,2,\cdots)$

$E[X_1] = V[X_1] = 2$

X_2 : 2時間あたりの来客数, r.v.

$X_2 \sim Po(4)$

$p_2(x) = e^{-4} \dfrac{4^x}{x!} \quad (x=0,1,2,\cdots)$

X_3 : 3時間あたりの来客数, r.v.

$X_3 \sim Po(6)$

$p_3(x) = e^{-6} \dfrac{6^x}{x!} \quad (x=0,1,2,\cdots)$

(1) 1時間に一人も客が来ない確率は?

$$P_1(0) = e^{-2}\frac{2^0}{0!}$$

$$= e^{-2} \fallingdotseq 0.1353$$

約13.5%

(2) 2時間に一人も客が来ない確率は?

$$P_2(0) = e^{-4}\cdot\frac{4^0}{0!}$$

$$= e^{-4} \fallingdotseq 0.01831$$

約1.8%

(3) 3時間に一人も客が来ない確率は?

$$P_3(0) = e^{-6}\frac{6^0}{0!}$$

$$= e^{-6} \fallingdotseq 0.002478$$

約0.25%

Th (正規分布の再生性)

$$X \sim N(\mu_X, \sigma_X^2)$$
$$Y \sim N(\mu_Y, \sigma_Y^2)$$
$$X, Y : \text{independent}$$
$$\Rightarrow X+Y \sim N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$$

Proof.

We will prove that

$$M_{X+Y}(t) = e^{(\mu_X+\mu_Y)t + \frac{1}{2}(\sigma_X^2+\sigma_Y^2)t^2}.$$

$$\begin{aligned}
M_{X+Y}(t) &= M_X(t)\, M_Y(t) \\
&= e^{\mu_X t + \frac{1}{2}\sigma_X^2 t^2} \cdot e^{\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2} \\
&= e^{(\mu_X+\mu_Y)t + \frac{1}{2}(\sigma_X^2+\sigma_Y^2)t^2}.
\end{aligned}$$

//

$$X \sim N(\mu, \sigma^2)$$
$$\Leftrightarrow M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Th

$X \sim N(\mu, \sigma^2)$

$a \in \mathbb{R}$

$\Rightarrow aX \sim N(a\mu, a^2\sigma^2)$

Proof

We will prove that

$$M_{aX}(t) = e^{(a\mu)t + \frac{1}{2}(a^2\sigma^2)t^2}.$$

We have that

$$M_{aX}(t) = M_X(at)$$
$$= e^{\mu(at) + \frac{1}{2}\sigma^2(at)^2}$$
$$= e^{(a\mu)t + \frac{1}{2}(a^2\sigma^2)t^2}.$$

//

Th

X : r.v.

$M_X(t) = E(e^{tX}]$

$a \in \mathbb{R}$

$M_{aX}(t) = E[e^{t\cdot(aX)}]$

the MGF for aX

$\Rightarrow M_{aX}(t) = M_X(at)$

Th

$X \sim N(\mu_X, \sigma_X^2)$

$Y \sim N(\mu_Y, \sigma_Y^2)$

X, Y : independent

$a, b \in \mathbb{R}$: const.

$\Rightarrow aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$

2つの場合

Proof

It holds that

$$\begin{cases} aX \sim N(a\mu_X, a^2\sigma_X^2) \\ bY \sim N(b\mu_Y, b^2\sigma_Y^2) \end{cases}.$$

Thus,

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2).$$

//

Cor

$X, Y \sim$ iid $N(\mu, \sigma^2)$

$$\Rightarrow \frac{1}{2}(X+Y) \sim N\left(\mu, \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2\right)$$

$$= N\left(\mu, \frac{\sigma^2}{2}\right)$$

※厳密には、上の Th で、

$$\boxed{\begin{array}{l} X, Y : \text{independent} \\ \Rightarrow aX, bY : \text{independent} \end{array}}$$

を証明しておく必要がある。

3つの場合

Th

$X \sim N(\mu_X, \sigma_X^2)$

$Y \sim N(\mu_Y, \sigma_Y^2)$

$Z \sim N(\mu_Z, \sigma_Z^2)$

X, Y, Z : independent

$a, b, c \in \mathbb{R}$: const.

$\Rightarrow aX+bY+cZ \sim N(a\mu_X+b\mu_Y+c\mu_Z,\ a^2\sigma_X^2+b^2\sigma_Y^2+c^2\sigma_Z^2)$

Proof

It holds that

$$\begin{cases} aX+bY \sim N(a\mu_X+b\mu_Y,\ a^2\sigma_X^2+b^2\sigma_Y^2). \\ cZ \sim N(c\mu_Z,\ c^2\sigma_Z^2). \end{cases}$$

Thus,

$$\begin{aligned} &aX+bY+cZ \\ &= (aX+bY)+cZ \\ &\sim N(a\mu_X+b\mu_Y+c\mu_Z,\ a^2\sigma_X^2+b^2\sigma_Y^2+c^2\sigma_Z^2). \end{aligned}$$

//

Cor

$X, Y, Z \sim \text{iid } N(\mu, \sigma^2)$

$$\begin{aligned} \Rightarrow \frac{1}{3}(X+Y+Z) &\sim N\left(\mu,\ \frac{1}{9}\sigma^2+\frac{1}{9}\sigma^2+\frac{1}{9}\sigma^2\right) \\ &= N\left(\mu,\ \frac{\sigma^2}{3}\right) \end{aligned}$$

前の定理は次のようにまとめることができる。

Th

$X_i \sim N(\mu_i, \sigma_i^2)$ $(i=1,2,\cdots,n)$

independent

$c_i \in \mathbb{R}$: const. $(i=1,2,\cdots,n)$

$$\Rightarrow \sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

Cor

$X_i \sim \text{iid } N(\mu, \sigma^2)$ $(i=1,2,\cdots,n)$

$$\Rightarrow S_n \equiv \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

Proof

$$\begin{cases} \mu_i = \mu \\ \sigma_i = \sigma \\ c_i = 1 \end{cases}$$ のケースなので（上の定理の）

$$S_n \equiv \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu, \sum_{i=1}^{n} \sigma^2\right)$$

$$= N(n\mu, n\sigma^2).$$

//

Cor

$$X_1, \cdots, X_n \sim iid\ N(\mu, \sigma^2)$$

$$\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Rightarrow \overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Proof

$$\begin{cases} \mu_i = \mu \\ \sigma_i = \sigma \\ c_i = \frac{1}{n} \end{cases}$$ のケースなので（前ページ）

$$\overline{X}_n \sim N\left(\mu, \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\right)$$

$$= N\left(\mu, \frac{1}{n^2} n\sigma^2\right)$$

$$= N\left(\mu, \frac{\sigma^2}{n}\right).$$

//

$N(\mu, \sigma^2)$

$$X_1, \cdots, X_n \sim iid\ N(\mu, \sigma^2)$$

$$\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$$

このとき $\overline{X}_n$ も正規分布に従うことがわかる。

$$N\left(\mu, \frac{\sigma^2}{n}\right)$$

Th

$X \sim N(\mu, \sigma^2)$

$a, b \in \mathbb{R}$: const.

$\Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$

Proof

We will prove that

$$M_{aX+b}(t) = e^{(a\mu+b)t + \frac{1}{2}(a^2\sigma^2)t^2}.$$

It holds that

$$\begin{aligned}
M_{aX+b}(t) &= E[e^{t(aX+b)}] \\
&= E[e^{atX} \cdot e^{bt}] \\
&= e^{bt} E[e^{atX}] \\
&= e^{bt} M_X(at) \\
&= e^{bt} \cdot e^{\mu \cdot (at) + \frac{1}{2}\sigma^2 (at)^2} \\
&= e^{bt} \cdot e^{(a\mu)t + \frac{1}{2}(a^2\sigma^2)t^2} \\
&= e^{(a\mu+b)t + \frac{1}{2}(a^2\sigma^2)t^2}.
\end{aligned}$$

//

※ 中心極限定理の変形を導くときなどに使う。

Cor

$X \sim N(\mu, \sigma^2)$

$\sigma > 0$

$$Z \equiv \frac{X-\mu}{\sigma}$$

$\Rightarrow Z \sim N(0,1)$

Proof

It holds that

$$Z \equiv \frac{1}{\sigma} X - \frac{\mu}{\sigma}.$$

Since $X \sim N(\mu, \sigma^2)$, we obtain that

$$Z \sim N\left(\frac{1}{\sigma} E(X) - \frac{\mu}{\sigma}, \frac{1}{\sigma^2} V[X]\right)$$

$$= N\left(\frac{1}{\sigma}\mu - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2\right)$$

$$= N(0,1).$$

//

例.

国民の体重
$N(60, 10^2)$

ある船の乗船客の体重
$X_i \sim iid\ N(60, 10^2)$

$S_n \equiv \sum_{i=1}^{n} X_i$　n人乗ったときの乗船客の総体重

$\frac{S_n}{n} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$　〃　〃　平均体重

$\frac{S_n}{n} \sim N\left(60, \frac{100}{n}\right)$

$$\boxed{(1)\ E[S_n] = ?,\quad E\left[\frac{S_n}{n}\right] = ?}$$

$$\begin{aligned} E[S_n] &\equiv E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n] \\ &= \underline{60n\ \mathrm{kg}} \end{aligned}$$

期待値の線型性

$$\begin{aligned} E\left[\frac{S_n}{n}\right] &= \frac{1}{n} E[S_n] \\ &= \underline{60\ \mathrm{kg}} \end{aligned}$$

期待値の線型性

$$\boxed{(2)\ V[S_n] = ?,\quad V\left[\frac{S_n}{n}\right] = ?}$$

$$\begin{aligned} V[S_n] &= V[X_1 + \cdots + X_n] \\ &= V[X_1] + \cdots + V[X_n] \\ &= \underline{100n} \end{aligned}$$

$X_1, \cdots, X_n$: independent

nとともに大きくなる.

$$V\left[\frac{S_n}{n}\right] = \frac{1}{n^2} V[S_n] = \underline{\frac{100}{n}}$$

nとともに小さくなる.
大きい船の方が
ある意味"安定する".

(3) 100人乗りの船で乗船客の平均体重が62kgをこえる確率はいくらか？

$X_i \sim iid\ N(60, 10^2)$ なので正規分布の再生性より

$$S_n \sim N(60n, 100n)$$

$$\frac{S_n}{n} \sim N(60, 100/n),$$

今、$n=100$ のケースなので

$$S_{100} \sim N(6000, 100^2)$$

$$\frac{S_{100}}{100} \sim N(60, 1) \quad \text{である。}$$

$S_{100}/100$ を標準化した r.v. を z とおく：

$$z \equiv \frac{S_{100}/100 - 60}{1}$$

すると $z \sim N(0,1)$ である。

よって、

$$P\left(\frac{S_{100}}{100} > 62\right)$$

$$= P\left(\frac{S_{100}}{100} - 60 > 62 - 60\right)$$

$$= P\left(\frac{S_{100}/100 - 60}{1} > \frac{62-60}{1}\right)$$

$$= P(z > 2) = 0.0228$$

∴ 約2.3% である。

分布の再生性

問題1. 二項分布、ポワソン分布、正規分布のモーメント母関数は、それぞれ以下のとおりである。

$$\text{二項分布}\quad X \sim B(n,p) \Rightarrow M(t) = [pe^t + (1-p)]^n$$

$$\text{ポワソン分布}\quad X \sim Po(\lambda) \Rightarrow M(t) = e^{\lambda(e^t-1)}$$

$$\text{正規分布}\quad X \sim N(\mu,\sigma^2) \Rightarrow M(t) = e^{\mu t+\frac{1}{2}\sigma^2 t^2}$$

モーメント母関数を用いて、これらの分布の再生性を証明しなさい。

問題2. $X \sim N(\mu,\sigma^2)$、aを定数とする。このとき、$aX \sim N(a\mu, a^2\sigma^2)$を示しなさい。

問題3. 確率変数Xについて、$X \sim N(\mu,\sigma^2)$とする。以下の推論はどこか間違っている。どこが間違っているか？ また、どう修正すれば良いか？

『$2X = X + X$なので、正規分布の再生性により$2X \sim N(\mu+\mu,\ \sigma^2+\sigma^2)$、よって、$2X \sim N(2\mu, 2\sigma^2)$である。』

問題4. $X \sim N(\mu_X,\sigma_X^2)$、$Y \sim N(\mu_Y,\sigma_Y^2)$、$a,b$を定数とする。このとき、$aX + bY \sim N(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2)$を示しなさい。

問題5. $X_1,\cdots,X_n \sim iidN(\mu,\sigma^2)$とする。このとき、以下を示しなさい。

(1) $S_n \equiv \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$

(2) $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

問題6.

(1) $X \sim N(\mu_X,\sigma_X^2)$、$a,b$を定数とする。このとき、$aX + b \sim N(a\mu_X + b,\ a^2\sigma_X^2)$を示しなさい。

(2) $X \sim N(\mu,\sigma^2)$とする。ただし、$\sigma > 0$を仮定する。このとき、$Z \equiv \frac{X-\mu}{\sigma} \sim N(0,1)$を示しなさい。

問題7. あなたは製薬会社に勤めており、2日続けて病院に営業に来た。待合室には風邪の患者が、昨日は10人、今日は7人待っていた。今、インフルエンザが流行っており、風邪の患者のうち5人に一人はインフルエンザだという。

(1) 昨日、インフルエンザ患者と接触した確率を求めなさい。

(2) 今日、インフルエンザ患者と接触した確率を求めなさい。

(3) この二日間の間に、この病院でインフルエンザ患者と接触した確率を求めなさい。

問題8. あなたは航空会社に勤めている。乗客の体重を表す確率変数Xは、平均$60kg$、分散64の正規分布に従うとする。$S_n = \sum_{i=1}^n X_i$とおくと、これは乗客数がn人のときの総体重、$S_n/n = \frac{1}{n}\sum_{i=1}^n X_i$は乗客数が$n$人のときの平均体重である。

(1) S_nとS_n/nの期待値を求めなさい。

(2) S_nとS_n/nの分散を求めなさい。

(3) 定員100人ジャンボジェット機が満席とする。乗客の平均体重が$62kg$を越える確率を求めなさい。