A study memo on popular probability distributions

Contents

1 General topics on random variables

By the definition of independence, the following clearly holds.

Proposition 1.1. *Let*

(S1) $(\S_i, \mathscr{S}, P_i)$ $(i = 1, 2, ..., N)$ *is a sequence of probability spaces.*

(S2) $(\Omega, \mathscr{F}, P)$ *is the probability spaces which is direct product of* $(\S_i, \mathscr{S}, P_i)$ $(i = 1, 2, ..., N)$

(S3) X_i *is a random variable on* S_i $(i = 1, 2, ..., N)$.

(S3) We set $Y_i := X_i \circ \pi_i$ $(i = 1, 2, ..., N)$.

then $Y_1, ..., Y_N$ *is a sequence of independent random variables.*

The following clearly holds.

Proposition 1.2. *Let* P *is probability measure on* $(\Omega := \mathbb{N} \cup \{0\}, 2^\Omega)$. *Then* id_Ω *is random variable on* Ω *and* $id_\Omega \sim P$.

By Fubini's theorem(see [5]), the following two propositions clearly holds.

Proposition 1.3 (Marginal distribution)**.** *Let*

(S1) $(\Omega_i, \mathscr{F}_i, P_i)$ *is a probability spaces* $(i = 1, 2)$.

(A1) $P_1 \times P_2$ *has a density function* f_{P_1,P_2}.

Then for almost everywhere $x \in \mathbb{R}$, $f_{P_1,P_2}(x, \cdot)$ *is measurable and*

$$f_{P_1}(x) := \int_{\mathbb{R}} f_{P_1,P_2}(x, y) dP_2(y)$$

exists and f_{P_1} *is measurable and*

$$\int_{\mathbb{R}} f_{P_1}(x) dP_1(x) = 1$$

Proposition 1.4 (Conditional probability density function)**.** *Let*

(S1) $(\Omega_i, \mathscr{F}_i, P_i)$ *is a probability spaces* $(i = 1, 2)$.

(A1) $P_1 \times P_2$ *has a density function* $f_{X,Y}$.

(S2) $x \in \mathbb{R}$ *such that* $f_{X,Y}(x, \cdot)$ *is measurable and* $f_X(x) > 0$.

(S3) Set

$$f_{P_2|P_1(x)}(y) := \frac{f_{P_1,P_2}(x, y)}{f_{P_1}(x)} \ (y \in \mathbb{R})$$

We call $f_{P_2|P_1(x)}$ *the conditional probability density function of* P_2 *given the occurrence of the value* x *of* P_1.

Then

$$\int_{\mathbb{R}} f_{P_2|P_1(x)}(y) dP_2(y) = 1$$

The following definitions are based on [6].

Definition 1.1 (Probability model, True distribution, Prior probability)**.** *Let*

(A1) *Q is a probability borel measure on $\mathbb{R}^N$ and Q has the density function q. We call q a true distribution.*

(S1) *W is a Borel set of $\mathbb{R}^d$.*

(A2) *Φ is a probability borel measure on W Φ has the density function ϕ. We call ϕ a prior probability.*

(A3) *$Q \times \Phi$ has the densition function p.*

(S2) *set $p(\cdot_1|\cdot_2)$ by for $w \in W$ such that $p_Q(w) > 0$*

$$p(x|w) := p(x|\Phi(w)) \ (x \in \mathbb{R})$$

We call $p(\cdot_1|\cdot_2)$ the a probability model. Or, we denote $p(\cdot_1|\cdot_2)$ by $p(x|w)$.

2 Probability generating function

Definition 2.1 (Probability generating function)**.** *Let*

(S1) *$(\Omega = \mathbb{N} \cup 0, 2^\Omega, P)$ is a probability space.*

then we set

$$G_P(z) := \sum_{i=0}^{\infty} P(i) z^i \ (z \in \mathbb{C}) \tag{1}$$

Proposition 2.1. *The followings hold.*

(i) *Radius of convergence of $G_P(z)$ is not less than 1.*

(ii) *If $G_P = G_{P'}$ then $P = P'$.*

(iii) *If Y is a random variable on any probability space such that $Y \sim P$ then $G_P(z) = E(z^Y)$ for any $z \in D(0,1)$.*

(iii) *If Y_1, Y_2 is a random variable on any probability space such that Y_1, Y_2 are independent then $G_{P_{Y_1+Y_2}} = G_{P_{Y_1}} G_{P_{Y_2}}$.*

proof of (i). Because $0 \le P \le 1$, (i) holds. □

proof of (ii). By (i) and definition of G_P and G'_P, (ii) holds. □

proof of (iii). Let us fix any $z \in D(0,1)$. For any $N \in \mathbb{N}$,

$$\begin{aligned} E(z^Y) &= \sum_{i=0}^{N} \int_{\{Y=i\}} z^Y dQ + \int_{\{Y>N\}} z^Y dQ \\ &= \sum_{i=0}^{N} P(i) z^i + \int_{\{Y>N\}} z^Y dQ \end{aligned} \tag{2}$$

So

$$|E(z^Y) - \sum_{i=0}^{N} P(i) z^i| \le |\int_{Y>N} z^Y dQ| \le Q(\{Y > N\}) \tag{3}$$

Consequently (iii) holds. □

proof of (iv). It is enough to show (iv) by (iii). □

3 Bernoulli distribution

Definition 3.1 (Bernoulli distribution)**.** *We call a probability distribution P on $\{0,1\}$ the Bernoulli distribution if for some $p \in [0,1]$ $P(\{1\}) = p$ and $P(\{0\}) = 1-p$.*

Proposition 3.1 (Expectation and Variance of Bernoulli distribution)**.** *Let us assumel a probability distribution P on $\{0,1\}$ is the Bernoulli distribution with $P(\{1\}) = p$.*

(i) $E(P) = p$

(ii) $V(P) = p(1-p)$,

(i). It is trivial. □

(i). $V(P) = \int_{\{0,1\}} x^2 dP - E(P)^2 = \int_{\{0,1\}} x dP - p^2 = p - p^2 = p(1-p)$ □

4 Binomial distribution

Definition 4.1 (Binomial distribution)**.** *For some $p \in [0,1]$ and $n \in \mathbb{N}$ we call a probability distribution $B(n,p)$ on $\{0,1,...,n\}$ the Binomial distribution if $B(n,p)(\{i\}) = {}_nCip^i(1-p)^(n-i)$ $(i = 0,1,...,n)$.*

Clearly the following holds.

Proposition 4.1. *Let*

(S1) $(\Omega, \mathscr{F}, P)$ *is a probability space.*

(S2) $\{X_i\}_{i=1}^n$ *be independent random variables.*

(A1) The distribution of X_i is the Bernoulli distribution B with $B(\{1\}) = p$ $(\forall i)$.

then the distribution of $\Sigma_{i=1}^n X_i$ is $B(n,p)$.

By Proposition Proposition1.2 and Proposition1.1, Random variables like the one above exist.
$E(B(2,p)) = 1 \cdot {}_2C_1 p(1-p) + 2 \cdot {}_2C_2 p^2 = 2p + 0 \cdot p^2 = 2p$. $E_{B(2,p)}(x^2) = 2p + 2^2p^2 - 2p^2$. $E(B(3,p)) = 1 \cdot {}_3C_1 p(1-p)^2 + 2 \cdot {}_3C_2 p^2(1-p) + 3p^3 = 3p + 0 \cdot p^2 + 0 \cdot p^3 = 3p$. $E_{B(3,p)}(x^2) = 3p + 3^3p^2 - 3p^2 + 0 \circ p^3$. We can extend these fact to the following lemma and the following proposition.

Lemma 4.1.

(i) $\Sigma_{k=1}^l k {}_lC_k(-1)^k = 0$ $(\forall l \geq 2)$.

(ii) $\Sigma_{k=1}^l k^2 {}_lC_k(-1)^k = 0$ $(\forall l \geq 3)$.

(i). $L(x) := (1-x)^l = \Sigma_{k=1}^l {}_lC_k(-1)^k(-1)^k x^k$.
$L'(x) = l(1-x)^{l-1} = \Sigma_{k=1}^l k {}_lC_k(-1)^k(-1)^k x^{k-1}$.

So, if $l \geq 2$, then

$$\begin{aligned} 0 &= L'(1) \\ &= \Sigma_{k=1}^l k {}_lC_k(-1)^k(-1)^k \end{aligned} \quad (4)$$

□

(ii). $L(x) := (1-x)^l = \Sigma_{k=1}^{l}{}_lC_k(-1)^k(-1)^k x^k$.
$L''(x) = l(1-x)^{l-1} = \Sigma_{k=1}^{l} k(k-1){}_lC_k(-1)^k(-1)^k x^{k-2}$.

So, if $l \geq 3$, then

$$\begin{aligned} 0 &= L''(1) \\ &= \Sigma_{k=1}^{l} k(k-1){}_lC_k(-1)^k(-1)^k \\ &= \Sigma_{k=1}^{l} k^2{}_lC_k(-1)^k(-1)^k - \Sigma_{k=1}^{l} k{}_lC_k(-1)^k(-1)^k \end{aligned} \tag{5}$$

By (i), $\Sigma_{k=1}^{l} k{}_lC_k(-1)^k(-1)^k = 0$. So $\Sigma_{k=1}^{l} k^2{}_lC_k(-1)^k(-1)^k = 0$. □

Proposition 4.2 (Expectation and Variance of Binomial distribution)**.**

(i) $E(B(n,p)) = np$

(ii) $V(B(n,p)) = np(1-p)$

proof1 of (i). Let us take $\{X_i\}_{\{}i = 1, 2., , , , n\}$ in Proposition4.1. $E(B(n,p)) = E(\Sigma_{i=1}^{n} X_i) = \Sigma_{i=1}^{n} E(X_i) = np$ □

proof1 of (ii). Let us take $\{X_i\}_{\{}i = 1, 2., , , , n\}$ in Proposition4.1. $V(B(n,p)) = \Sigma_{i=1}^{n} V(X_i) = np(1-p)$ □

proof2 of (i).

$$\begin{aligned} E(B(n,p)) &= \sum_{k=1}^{n} k{}_nC_k p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{l} k{}_nC_k p^k \sum_{i=0}^{n-k} {}_{n-k}C_i(-1)^i p^i \\ &= \sum_{l=1}^{n} \sum_{k=1,2\ldots,l,\ i=0,1,\ldots,n-k,\ k+i=l} k{}_nC_k p^k{}_{n-k}C_i(-1)^i p^i \\ &= \sum_{l=1}^{n} p^l \sum_{k=1,2\ldots,l,\ i=0,1,\ldots,n-k,\ k+i=l} k{}_nC_k{}_{n-k}C_i(-1)^i \\ &= \sum_{l=1}^{n} p^l \sum_{k=1}^{l} k{}_nC_k \cdot {}_{n-k}C_{l-k}(-1)^{l-k} \\ &= \sum_{l=1}^{n} (-1)^l p^l \sum_{k=1}^{l} k{}_nC_k \cdot {}_{n-k}C_{l-k}(-1)^k \\ &= \sum_{l=1}^{n} (-1)^l p^l \sum_{k=1}^{l} k\frac{{}_nP_l}{k!(l-k)!}(-1)^k \\ &= \sum_{l=1}^{n} (-1)^l p^l \sum_{k=1}^{l} k\frac{{}_nC_l \cdot l!}{k!(l-k)!}(-1)^k \\ &= \sum_{l=1}^{n} (-1)^l p^l{}_nC_l \sum_{k=1}^{l} k\frac{l!}{k!(l-k)!}(-1)^k \\ &= \sum_{l=1}^{n} (-1)^l p^l{}_nC_l \sum_{k=1}^{l} k{}_lC_k(-1)^k \end{aligned} \tag{6}$$

By Lemma4.1, for any $l \geq 2$, $\sum_{k=1}^{l} k_l C_k (-1)^k = 0$. So $E(B(n,p)) = np$. □

proof2 of (ii). By the proof2 of (ii),

$$E_{B(n,p)}(x^2) = \sum_{l=1}^{n} (-1)^l p^l {}_n C_l \sum_{k=1}^{l} k^2 {}_l C_k (-1)^k \tag{7}$$

By Lemma4.1, for any $l \geq 3$, $\sum_{k=1}^{l} k^2 {}_l C_k (-1)^k = 0$.

So $E_{B(n,p)}(x^2) = \sum_{l=1}^{2} (-1)^l p^l {}_n C_l \sum_{k=1}^{l} k^2 {}_l C_k (-1)^k = np(1-p) + n^2 p^2$. By (i), $V(B(n,p)) = E_{B(n,p)}(x^2) - E(B(n,p))^2 = np(1-p)$. □

5 Geometric distribution

Definition 5.1 (Geometric distribution)**.** *Let $p \in (0,1)$.*

$$P(k) := (1-p)^{k-1} p \ (k = 1, 2, ...) \tag{8}$$

We call P is Geometric distribution with p

Clearly P is a probability measure on $\{1, 2, ..., n, ...\}$.

Proposition 5.1. *Let P is Geometric distribution with p. Then*

$$G_P(z) = \frac{pz}{1-(1-p)z} \tag{9}$$

Proof.

$$\begin{aligned} G_P(z) &= \sum_{k=1} (1-p)^{k-1} p z^k \\ &= pz \sum_{k=1} (1-p)^{k-1} p z^{k-1} \\ &= pz \frac{1}{1-(1-p)z} \end{aligned} \tag{10}$$

□

Proposition 5.2. *Let P is Geometric distribution with p. Then*

$$E(P) = \frac{1}{p} \tag{11}$$

and

$$V(P) = \frac{1-p}{p^2} \tag{12}$$

proof1 of (11).

$$G'_P(z) \quad = \quad \frac{p(1-(1-p)z)+pz(1-p)}{(1-(1-p)z)^2}$$

So

$$\begin{aligned} E(P) &= G'_P(1) = \frac{p(1-(1-p)1)+p1(1-p)}{(1-(1-p)1)^2} \\ &= \frac{p^2+p-p^2)}{p^2} = \frac{1}{p} \end{aligned} \tag{13}$$

□

proof2 of (11).

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \tag{14}$$

By calculating the derivative,

$$\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1} \tag{15}$$

So

$$E(P) = p\sum_{k=1}^{\infty} k(1-(1-p))^{k-1} = p\frac{1}{(1-(1-p))^2} = \frac{1}{p} \tag{16}$$

□

proof of (12). By calculating the derivative of (17),

$$\frac{2}{(1-x)^3} = \sum_{k=2}^{\infty} k(k-1)x^{k-2} \tag{17}$$

So

$$\begin{aligned} E_P(x(x-1)) &= p\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1} \\ &= p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= p(1-p)\frac{2}{(p^3} = \frac{2(1-p)}{p^2} \end{aligned} \tag{18}$$

$$V(P) = E_P(x(x-1)) + E_P(x) - E_P(x)^2 = \frac{2(1-p)}{p^2} + \frac{p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} \tag{19}$$

□

6 Negative binomial distribution

Definition 6.1 (Negative binomial distribution). *We call a probability distribution P on $\{1, 2, ...\}$ the Negative binomial distribution if for some $p \in [0, 1]$ $P(\{k\}) = {}_{r+k-2}C_{r-1}(1-p)^{k-1}p^{r-1}$. We denote this distribution by $NB(r, p)$.*

Proposition 6.1.

$$G_{NB(r,p)}(z) = \frac{p^r z}{(1-(1-p)z)^r} \tag{20}$$

Proof. Because

$$\frac{1}{1-z} = \sum_{i=0}^{\infty} z^i \tag{21}$$

the following holds by $r-1$ times derivative.

$$\frac{(r-1)!}{(1-z)^r} = \sum_{i=r-1}^{\infty} i(i-1)...(i-r+2)z^i \tag{22}$$

□

Proposition 6.2. *Let $X_1, ..., X_r$ are independent random variables and for any i P_{X_i} is the geometric distribution. Then the distribution of $\sum_{i=1}^{r} X_i - (r-1)$ is $N(r, p)$.*

References

[1] Tadahisa Funaki, Probability Theory(in Japanese), ISBN-13 978-4254116007.

[2] Shinichi Kotani, Measure and Probability(in Japanese), ISBN4-00-010634-1.

[3] A study memo on a proof of the central limit theorem https://osmanthus.work/?p=607

[4] Tatsuya Kubota, Foundations of modern mathematical statistics(in Japanese), ISBN978-4-320-11166-0.

[5] Seizo Ito, Introduction to Lebesgue Integral(in Japanese), ISBN978-4785313180.

[6] Sumio Watanabe, Theory and Methods of Bayesian Statistics(in Japanese), ISBN978-4339024623.