Домашняя работа

Митин Арсений

Домашнее задание 1

Выбор распределений

Выбранные распределения:

- Дискретное: *гипергеометрическое*
- Непрерывное: *нормальное*

Описание основных характеристик распределений

Гипергеометрическое распределение

Гипергеометрическое распределение - дискретное распределение, описывающее вероятность события, при котором ровно k из n случайно выбранных элементов окажутся *помеченными*, при этом выборка осуществляется из множества мощности N, в котором присутствует m помеченных элементов. Считается, что каждый из элементов может быть выбран с одинаковой вероятностью $\frac{1}{N}$. Запишем это формально:

$$N \in \mathbb{N},\ m \in \overline{0,N},\ n \in \overline{0,N},$$
$$k \in \overline{0,n}$$

Тогда $HG(D,N,n)$ описывает вероятность события, при котором ровно k из n элементов выборки окажутся *помеченными*:

$$\{\xi \sim HG(N,m,n)\}$$
$$\Updownarrow$$
$$\left\{\mathbb{P}\left(\,\xi = k\,\right) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}\right\}$$

Математическое ожидание гипергеометрического распределения По определению, математическое ожидание случайной величины ξ – это ее $1^{й}$ начальный момент. Для начала, найдем $k^{й}$ начальный момент для ξ:

$$\mathbb{E}\left[\,\xi^r\,\right] = \sum_{k=0}^{n} k^r \cdot \mathbb{P}\left(\,\xi = k\,\right) = \sum_{k=0}^{n} k^r \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

Можем считать, что сумма берется при k от 1 до n,так как слагаемое при $k = 0$ будет равно 0. Заметим, что

$$\begin{aligned} k\binom{m}{k} &= k\frac{m!}{k!(m-k)!} = \\ &= k\frac{m \cdot (m-1)!}{k \cdot (k-1)! \cdot (m-k)!} = \\ &= m\frac{(m-1)!}{(k-1)! \cdot (m-1-(k-1))!} = \\ &= m\binom{m-1}{k-1} \end{aligned}$$

и, как следствие,

$$\binom{N}{n} = \frac{1}{n} \cdot n \cdot \binom{N}{n} = \frac{1}{n}N\binom{N-1}{n-1}$$

Подставим TODO и TODO в TODO:

$$\mathbb{E}\left[\,\xi^r\,\right] = \frac{n \cdot m}{N}\sum_{k=1}^{r-1} \frac{\binom{m-1}{k-1}\binom{N-m}{n-k}}{\binom{N-1}{n-1}}$$

Положим $j := k - 1$ и изменим индекс суммирования с на $j = \overline{0, n-1}$. Заметим, что $n - k = n - (j + 1) = (n - 1) - j$ и $N - m = (N - 1) - (m - 1)$:

$$\mathbb{E}\left[\,\xi^r\,\right] = \frac{n \cdot m}{N}\sum_{j=0}^{n-1} (j+1)^{r-1} \frac{\binom{m-1}{j}\binom{(N-1)-(m-1)}{(n-1)-j}}{\binom{N-1}{n-1}}$$

Заметим, что выделенная часть выражения может быть записана, как $\mathbb{E}\left[\,(\theta+1)^{r-1}\,\right]$, где $\theta \sim HG(N-1, m-1, n-1)$. Следовательно,

$$\mathbb{E}\left[\,\xi^r\,\right] = \frac{n \cdot m}{N}\mathbb{E}\left[\,(\theta+1)^{r-1}\,\right]$$

Таким образом,

$$\boxed{\mathbb{E}\left[\,\xi\,\right] = \frac{n \cdot m}{N}}$$

Дисперсия гипергеометрического распределения По определению дисперсии,

$$\text{Var}\left[\,\xi\,\right] = \mathbb{E}\left[\,(\xi - \mathbb{E}\left[\,\xi\,\right])^2\,\right] = \\ = \mathbb{E}\left[\,\xi^2\,\right] - (\mathbb{E}\left[\,xi\,\right])^2$$

Выведем $2^{\text{й}}$ начальный момент из TODO:

$$\mathbb{E}\left[\,\xi^2\,\right] = \frac{n \cdot m}{N}\mathbb{E}\left[\,\theta + 1\,\right] = \frac{n \cdot m}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1\right)$$

Подставим TODO и TODO в TODO:

$$\begin{aligned}\text{Var}\left[\,\xi\,\right] = \mathbb{E}\left[\,\xi^2\,\right] - (\mathbb{E}\left[\,\xi\,\right])^2 = \\ &= \frac{n \cdot m}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1\right) - \left(\frac{n \cdot m}{N}\right)^2 = \\ &= \frac{n \cdot m}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{n \cdot m}{N}\right)\end{aligned}$$

Таким образом,

$$\boxed{\mathbb{E}\left[\,\xi\,\right] = \frac{n \cdot m}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{n \cdot m}{N}\right)}$$

Производящая функция гипергеометрического распределения По определению производящей функции,

$$M_\xi(t) = \mathbb{E}\left[\,e^{t\xi}\,\right]$$

То есть,

$$M_\xi(t) = \sum_{k=0}^{n} e^{tk}\mathbb{P}\left(\,\xi = k\,\right) = \\ = \sum_{k=0}^{n} e^{tk}\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

TODO

Характеристическая функция гипергеометрического распределения TODO

Гистограмма вероятностей гипергеометрического распределения Построим гистограмму вероятностей для $k \in \overline{0, n}$:

Hist HG

Функция распределения гипергеометрического распределения По определению, функция распределения $F_\xi(k) = \mathbb{P}\left(\, \xi < k \,\right)$. Событие $\{\xi < k\} = \bigcup\limits_{i=0}^{k-1} \{\xi = i\}$. События $\{\xi = i\}$ $\forall i \in \overline{0, k-1}$ являются попарно несовместными. То есть $\forall i, j \in \overline{0, k-1} : i \neq j$ выполняется $\{\xi = i\} \cap \{\xi = j\} = \emptyset$. Из этого следует, что

$$\mathbb{P}\left(\, \xi < k \,\right) = \sum_{i=0}^{k-1} \mathbb{P}\left(\, \xi = i \,\right)$$

Подставим TODO в это выражение и получим:

$$F_\xi(k) = \sum_{i=0}^{k-1} \mathbb{P}\left(\, \xi = i \,\right) = \sum_{i=0}^{k-1} \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

Построим график этой функции, учитывая, что аргументом k должно быть натуральное число, не превосходящее n:

TODO

Нормальное распределение

Нормальное распределение - непрерывное распределение, описывающее поведение величины отклонения измеряемого значения x от истинного значения μ (которое является математическим ожиданием) и в рамках некоторого разброса σ (среднеквадратичного отклонения). Запишем это формально:

$$\{\eta \sim N(\mu, \sigma^2)\}$$
$$\Updownarrow$$
$$\left\{ \begin{matrix} F_\eta(x) = \mathbb{P}\left(\, \eta < x \,\right) = \displaystyle\int_{-\infty}^{x} f_\eta(x)dx, \\ \text{где} f_\eta(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \end{matrix} \right\}$$

$f_\eta(x)$ называется плотностью вероятности.

Математическое ожидание нормального распределения Найдем математическое ожидание $\eta \sim N(\mu, \sigma^2)$:

$$\begin{aligned}\mathbb{E}\left[\,\eta\,\right] &= \int_{-\infty}^{+\infty} x \cdot f_\eta(x) dx = \\ &= \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx\end{aligned}$$

Сделаем замену $t = \frac{x-\mu}{\sqrt{2}\sigma}$:

$$\begin{aligned}\mathbb{E}\left[\,\eta\,\right] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma\sqrt{2}t + \mu) e^{-t^2} d\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) = \\ &= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t e^{-t^2} dt + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \\ &= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \left(\int_{-\infty}^{0} t e^{-t^2} dt - \int_{-\infty}^{0} t e^{-t^2} dt \right) + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \\ &= \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt\end{aligned}$$

Заметим, что получившееся выражение содержит интеграл, который может быть сведен к интегралу Эйлера-Пуассона:

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = 2\int_{0}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$$

Таким образом,

$$\boxed{\mathbb{E}\left[\,\eta\,\right] = \mu}$$

Дисперсия нормального распределения Подставим TODO в определение дисперсии TODO:

$$\begin{aligned}
\text{Var}\left[\,\eta\,\right] = \mathbb{E}\left[\,(\eta-\mu)^2\,\right] = \\
&= \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f_\eta(x)dx = \\
&= \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\end{aligned}$$

Сделаем ту же замену переменной $t = \frac{x-\mu}{\sqrt{2}\sigma}$, тогда $x = t\sqrt{2}\sigma + \mu$ и:

$$\begin{aligned}
\text{Var}\left[\,\eta\,\right] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sqrt{2}\sigma)^2 t^2 e^{-t^2} d(t\sqrt{2}\sigma + \mu) = \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt
\end{aligned}$$

Проинтегрируем по частям:

$$\begin{aligned}
\text{Var}\left[\,\eta\,\right] &= \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t2te^{-t^2} dt = \\
&= \frac{\sigma^2}{\sqrt{\pi}} \left(-te^{-t^2}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-t^2} dt \right)
\end{aligned}$$

Здесь снова появляется интеграл Эйлера-Пуассона и, в итоге, получаем:

$$\boxed{\text{Var}\left[\,\eta\,\right] = \sigma^2}$$

То есть, σ является среднеквадратичным отклонением.