دانشکده‌ی علوم ریاضی

احتمال و کاربردآن ۱۹ فروردین ۹۳

جلسه‌ی ۱۲: واریانس و متغیرهای تصادفی برنولی، دوجمله‌ای، پواسون

مدرّس: دکتر شهرام خزائی *نگارنده: آرمیتا خواجه‌نصیری*

۱ مروری بر فرمول‌های متغیر تصادفی

همان‌گونه که در جلسه قبل مشاهده کردیم، فرمول‌های زیر برای محاسبه متوسط یا امید ریاضی[1] یک متغیر تصادفی گسسته مانند X و تابعی مانند g از آن به‌کار برده می‌شود.

$$\mathrm{E}[X] = \sum_i x_i p_X(x_i)$$

$$\mathrm{E}[g(X)] = \sum_i g(x_i) p_X(x_i)$$

قضیه ۱ *فرض کنید $g_1, g_2, ... g_n$ توابعی از $\mathbb{R} \to \mathbb{R}$ و X یک متغیر تصادفی باشد. در این صورت برای محاسبه مقدار امید ریاضی خواهیم داشت:*

$$E[g_1(X) + g_2(X) + ... + g_n(X)] = E[g_1(X)] + E[g_2(X)] + ... E[g_n(X)]$$

مثال ۱ *عبارت زیر را ساده تر کنید.*

$$\mathrm{E}[X^3 - 2X^2 + e^X \cos(X) - 5\sin^2(X)]$$

پاسخ

$$\mathrm{E}[X^3 - 2X^2 + e^X \cos(X) - 5\sin^2(X)] = \mathrm{E}[X^3] - 2\mathrm{E}[X^2] + \mathrm{E}[e^X \cos(X)] - 5\mathrm{E}[\sin^2(X)]$$

۲ واریانس متغیر تصادفی

اگر یک متغیر تصادفی داشته باشیم، گرچه مقدار امید ریاضی آن مانند مرکز ثقل در فیزیک عمل کرده و متوسط وزنی مقادیر را می‌دهد، اما نمی‌تواند اطلاعاتی در مورد پراکندگی مقادیر ارائه دهد. برای مثال نمره دانش‌آموزی را در دو درس ریاضی و ادبیات در نظر بگیرید. اگر در درس ریاضی ۱۵ و در درس ادبیات

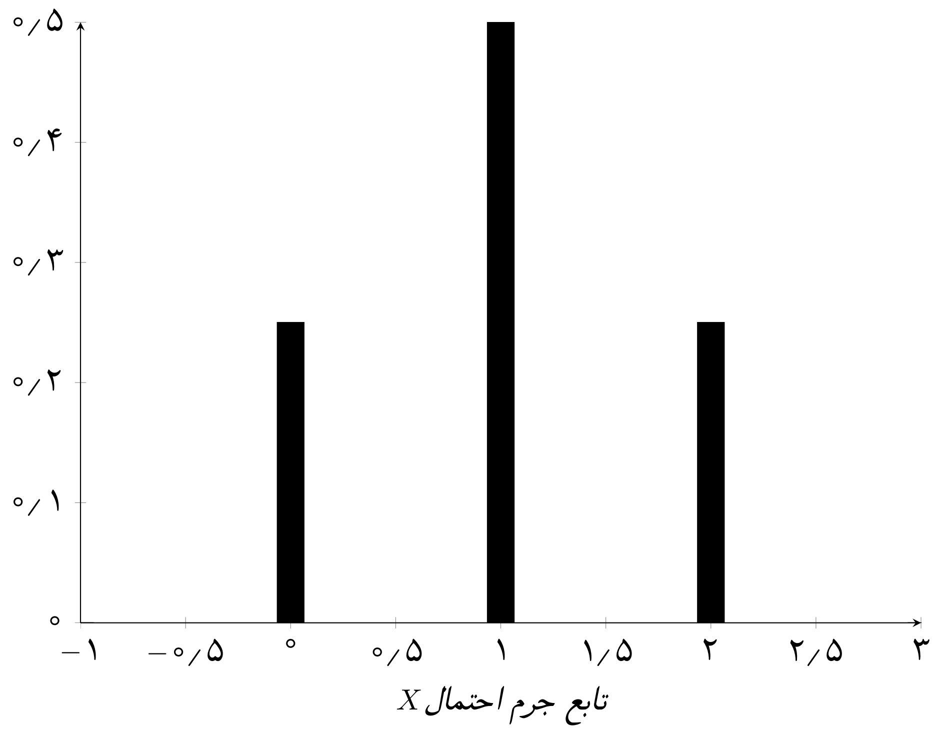
[1] Expected value

نمره ۱۸ را کسب کرده باشد، پاسخ به این سوال که در کدام درس بهتر عمل کرده بدون دانستن پراکندگی نمرات بقیه دانش‌آموزان حول میانگین ممکن نیست. پس برای گرفتن یک نتیجه درست نمی‌توانیم فقط به متوسط قانع باشیم و نیاز به معیارهای دیگری مثل واریانس داریم.

تعریف ۱ *واریانس[۲] متغیر تصادفی X به صورت زیر تعریف می‌شود:*

$$\sigma_X^۲ = \mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^۲]$$

مثال ۲ *فرض کنید تابع جرم احتمال متغیر تصادفی X و Y به صورت زیر داده شده باشد. واریانس این دو متغیر تصادفی را محاسبه کنید.*

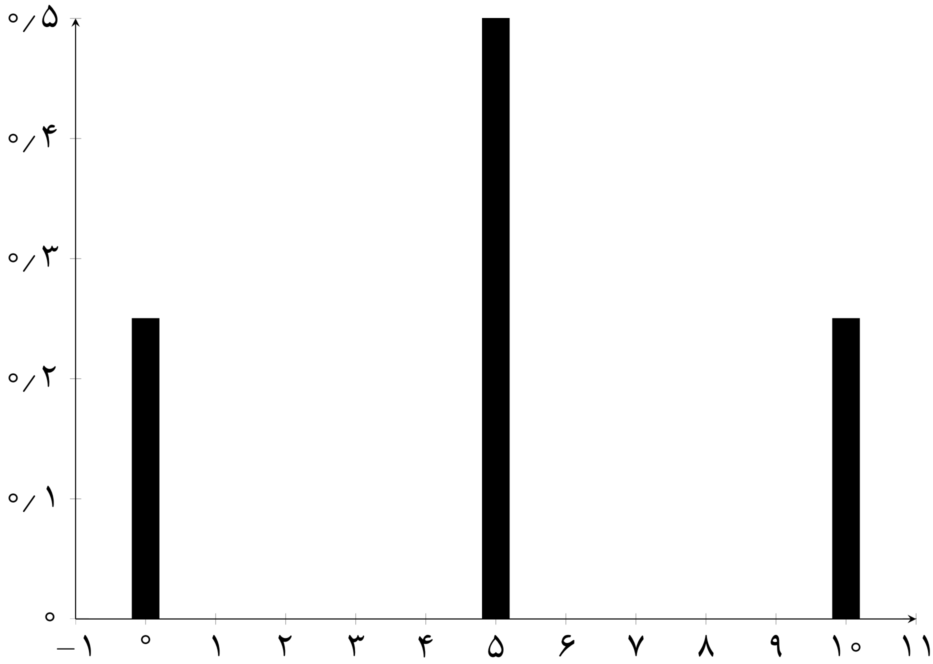


پاسخ *برای متغیر تصادفی X داریم:*

$$\mathrm{E}[X] = ۰ \times \frac{۱}{۴} + ۱ \times \frac{۱}{۲} + ۲ \times \frac{۱}{۴} = ۱$$

$$\sigma_X^۲ = (۰ - ۱)^۲ \times \frac{۱}{۴} + (۱ - ۱)^۲ \times \frac{۱}{۲} + (۲ - ۱)^۲ \times \frac{۱}{۴} = ۰/۵$$

[۲] Variance



تابع جرم احتمال Y

پاسخ *برای متغیر تصادفی Y داریم:*

$$\mathrm{E}[Y] = ۰ \times \frac{۱}{۴} + ۵ \times \frac{۱}{۲} + ۱۰ \times \frac{۱}{۴} = ۵$$

$$\sigma_Y^۲ = (۰ - ۵)^۲ \times \frac{۱}{۴} + (۵ - ۵)^۲ \times \frac{۱}{۲} + (۱۰ - ۵)^۲ \times \frac{۱}{۴} = ۱۲٫۵$$

نکته ۱ *محاسبه واریانس از رابطه زیر معمولاً ساده تر است.*

$$\sigma_X^۲ = E[X^۲] - E^۲[X]$$

برهان. داریم:

$$\begin{aligned}
\sigma_X^۲ &= \mathrm{Var}(X) \\
&= \mathrm{E}[(X - \mathrm{E}[X])^۲] \\
&= \mathrm{E}[X^۲ - ۲\mathrm{E}[X]X + \mathrm{E}^۲[X]] \\
&= \mathrm{E}[X^۲] - ۲\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}^۲[X] \\
&= \mathrm{E}[X^۲] - \mathrm{E}^۲[X]
\end{aligned}$$

■

حال با استفاده از نکته ۱ می‌توان واریانس متغیرهای تصادفی X و Y در مثال ۲ را به صورت زیر ساده‌تر محاسبه کرد:
متغیر تصادفی X:

$$\mathrm{E}[X^۲] = ۰^۲ \times \frac{۱}{۴} + ۱^۲ \times \frac{۱}{۲} + ۲^۲ \times \frac{۱}{۴} = ۱٫۵$$

$$\sigma_X^۲ = \mathrm{E}[X^۲] - \mathrm{E}^۲[X] = ۱/۵ - ۱ = ۰/۵$$

متغیر تصادفی Y:

$$\mathrm{E}[Y^۲] = ۰^۲ \times \frac{۱}{۴} + ۵^۲ \times \frac{۱}{۲} + ۱۰^۲ \times \frac{۱}{۴} = ۳۷/۵$$

$$\sigma_Y^۲ = \mathrm{E}[Y^۲] - \mathrm{E}^۲[Y] = ۱۲/۵$$

قضیه ۲ *اگر α و β مقادیر ثابتی باشند، واریانس به صورت زیر تغییر می‌کند:*

$$\mathrm{Var}[\alpha X + \beta] = \alpha^۲ \mathrm{Var}[X]$$

نکته ۲ *فرض کنید متغیر تصادفی X با تابع جرم احتمال p_X موجود است و مقدار t را برای پیش‌بینی X انتخاب کرده‌ایم. آنگاه متغیر تصادفی $X - t$ بیانگر خطای تخمین است. یک معیار برای یک تخمین مناسب کمینه کردن تابعی از قدر مطلق خطا است. معیاری که معمولاً در نظر گرفته می‌شود، متوسط مجذور خطا، یعنی مقدار $f(t) = \mathrm{E}[(X-t)^۲]$، را کمینه می‌کند. داریم:*

$$f(t) = \mathrm{E}[(X-t)^۲] = \mathrm{E}[X^۲] - ۲t\mathrm{E}[X^۲] + t^۲$$

$$\frac{\mathrm{d}f}{\mathrm{d}x} = ۰ + (-۲\mathrm{E}[X]) + ۲t = ۰$$

$$\Rightarrow t = \mathrm{E}[X]$$

پس متوسط مجذور خطا برای $t = \mathrm{E}[X]$ کمینه می‌شود و این مقدار مینیمم برابر واریانس است. زیرا

$$\mathrm{E}[(X-t)^۲] = \mathrm{E}[(X - \mathrm{E}[X])^۲] = \mathrm{Var}(X) \ .$$

پس هرچه مقدار $\mathrm{Var}(X)$ کمتر باشد، $\mathrm{E}[X]$ تقریب بهتری با معیار کمترین‌مقدار متوسط خطا (MMSE[۳]) برای X است. جالب توجه است که تخمین با معیار MMSE خودُ متوسط خطا را نیز صفر می‌کند که ویژگی مطلوبی است. زیرا

$$\mathrm{E}[(X-t)] = \mathrm{E}[X] - t = \mathrm{E}[X] - \mathrm{E}[X] = ۰ \ .$$

۳ انحراف معیار و نرمال کردن

تعریف ۲ *جذر واریانس را انحراف معیار[۴] می‌نامیم و توجه می‌کنیم که جنس آن با جنس متغیر تصادفی یکسان است.*

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

تعریف ۳ *متغیر تصادفی نرمال شده X را به صورت زیر تعریف می‌کنیم:*

$$X^* = \frac{X - \mathrm{E}[X]}{\sigma_X}$$

[۳] Minimum Mean Square Error

[۴] Standard deviation

۱. ویژگی اول متغیر تصادفی نرمال شده:

$$\mathrm{E}[X^*] = ۰$$
$$E[X^*] = E(\frac{۱}{\sigma_X}X - \frac{E[X]}{\sigma_X}) = (\frac{E[X]}{\sigma_X}) - (\frac{E[X]}{\sigma_X}) = ۰$$

۲. ویژگی دوم متغیر تصادفی نرمال شده:

$$\mathrm{Var}(X^*) = ۱$$
$$\mathrm{Var}[X^*] = \mathrm{Var}(\frac{۱}{\sigma_X}X - \frac{\mathrm{E}[X]}{\sigma_X}) = \frac{۱}{\sigma_X^۲}\mathrm{Var}(X) = \frac{\sigma_X^۲}{\sigma_X^۲} = ۱$$

نرمال کردن عموما موقعی مفید است که می‌خواهیم دو یا چند متغیر تصادفی با توزیع‌های مختلف را با هم مقایسه کنیم. برای مثال نمره دانش آموزی در درس احتمال ۷۲ و در درس تاریخ ۸۵ است. در نگاه اول به نظر می‌رسد که او در درس تاریخ عملکرد بهتری دارد. هرچند این گمان ممکن است نادرست باشد. برای مثال فرض کنید مقدار متوسط همه نمره‌ها در درس تاریخ ۸۲ و انحراف معیار ۷ است. و همین مقادیر برای درس احتمال ۶۸ و ۴ است. اگر نمرات دانش آموز را به صورت نرمال تبدیل کنیم، صورت نرمال شده نمره درس احتمال و تاریخ او به ترتیب $۱ = \frac{۷۲-۶۸}{۴}$ و $۰/۴۳ = \frac{۸۵-۸۲}{۷}$ است. که نشانگر این است که نمره نرمال شده‌ی او در درس احتمال ۱ و در درس تاریخ ۰.۴۳ است. پس او در درس احتمال بهتر از تاریخ عمل کرده است. اگر توزیع متغیرهای تصادفی داده شده باشد، مقایسه بهتری می‌توان انجام داد.

۴ متغیر تصادفی برنولی

فرض کنید آزمایشی انجام گیرد که نتیجه آن را بتوان به یکی از دو حالت موفقیت و شکست تقسیم‌بندی نمود. وقتی که موفقیت حاصل می‌شود متغیر تصادفی برابر ۱ و وقتی شکست حاصل می‌شود آن را برابر صفر قرار می‌دهیم. به این متغیر تصادفی یک متغیر تصادفی برنولی[۵] می‌گویند (به یاد یکی از برنولی‌ها به‌نام جیمز برنولی).

تعریف ۴ *توزیع برنولی با پارامتر p با تابع جرم احتمال زیر تعریف می شود:*

$$p_X(۰) = \Pr[X = ۰] = ۱ - p$$
$$p_X(۱) = \Pr[X = ۱] = p$$

حال به محاسبه مقدار متوسط و واریانس آن می‌پردازیم:

$$\mathrm{E}[X] = ۰ \times (۱ - p) + ۱ \times p = p$$
$$\mathrm{E}[X^۲] = ۰^۲ \times (۱ - p) + ۱^۲ \times p = p$$

(دقت که که به یک‌باره هم می‌توانستم استدلال کنیم که $\mathrm{E}[X^۲] = p$ زیرا $X^۲ = X$ و لذا $X^۲$ نیز دارای توزیع برنولی با پارامتر p است.)

$$\sigma_X^۲ = \mathrm{E}[X^۲] - \mathrm{E}[X]^۲ = p - p^۲ = p(۱ - p)$$

[۵] THE BERNOULLI RANDOM VARIABLE

۵ متغیر تصادفی دوجمله‌ای

فرض کنید n آزمایش مستقل که هر کدام با احتمال p به موفقیت و با احتمال $۱-p$ به شکست منجر می‌شوند را انجام داده و X نشان دهنده تعداد موفقیت‌ها در n آزمایش در نظر گرفته شود. آنگاه X را یک متغیر تصادفی دوجمله‌ای[۶] با پارامتر (n,p) گویند. بنابراین متغیر تصادفی برنولی همان متغیر تصادفی دوجمله‌ای با پارامترهای $(۱,p)$ است. تابع جرم احتمال یک متغیر تصادفی دوجمله‌ای با پارامترهای (n,p) به صورت زیر است:

$$p_X(k) = \binom{n}{k} p^k (۱-p)^{n-k}$$

$$k = ۱, ۲, .., n$$

حال به بررسی خصوصیات این توزیع می‌پردازیم:

۱. متوسط (امید ریاضی)

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=۰}^{n} k\binom{n}{k} p^k (۱-p)^{n-k} \\
&= \sum_{k=۱}^{n} k\frac{n!}{k!(n-k)!} p^k (۱-p)^{n-k} \\
&= \sum_{k=۱}^{n} \frac{n(n-۱)!}{(k-۱)!(n-k)!} p^k (۱-p)^{n-k} \\
&= np\sum_{k=۱}^{n} \frac{(n-۱)!}{(k-۱)!(n-k)!} p^{k-۱} (۱-p)^{n-k} \\
&= np\sum_{j=۰}^{n-۱} \frac{(n-۱)!}{(j)!(n-۱-j)!} p^j (۱-p)^{n-۱-j} \\
&= np\sum_{j=۰}^{n-۱} \binom{n-۱}{j} p^j (۱-p)^{n-۱-j}
\end{aligned}$$

عبارت جلوی $\sum$ آخرین جمله خود یک توزیع دوجمله‌ای است و چون می‌دانیم جمع احتمال‌ها باید یک باشد پس کّل عبارت بعد از $\sum$ برابر یک است. بنابرین:

$$\mathrm{E}[X] = np$$

۲. واریانس می‌توان نشان داد:

$$\begin{aligned}
\mathrm{E}[X^۲] &= \sum_{k=۱}^{n} k^۲\binom{n}{k} p^k (۱-p)^{n-k} \\
&= np[(n-۱)p+۱]
\end{aligned}$$

پس:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[X^۲] - (\mathrm{E}[X])^۲ \\
&= n^۲p^۲ - np^۲ + np - (np)^۲ \\
&= np(۱-p)
\end{aligned}$$

بنابراین:

$$\mathrm{Var}(X) = np(۱-p)$$

[۶] THE BINOMINAL RANDOM VARIABLE

قضیه ۳ *اگر X یک متغیر تصادفی دو جمله‌ای با پارامترهای (n,p) باشد، وقتی k از ۰ به n افزایش می‌یابد، تابع جرم احتمال ابتدا صعودی یکنوا و سپس نزولی یکنوا است و بیشترین مقدار خود را زمانی انتخاب می‌کند که k بزرگترین عدد صحیح کوچکتر یا مساوی $(n+۱)p$ را اختیار کند.*

برهان. *از حسابان می‌دانیم وقتی یک تابع صعودی و سپس نزولی است، مقدار ماکسیمم در نقطه‌ای حاصل می‌شود که نوع یکنوایی تغییر کند. برای اثبات از نسبت $\Pr\{X=k\}/\Pr\{X=k-۱\}$ استفاده می‌کنیم و مقداری از k را محاسبه می‌کنیم که این نسبت بزرگتر یا کوچکتر از ۱ است:*

$$\begin{aligned}\frac{\Pr\{X=k\}}{\Pr\{X=k-۱\}} &= \frac{\frac{n!}{(n-k)!k!}p^k(۱-p)^{n-k}}{\frac{n!}{(n-k+۱)!(k-۱)!}p^{k-۱}(۱-p)^{n-k+۱}}\\ &= \frac{(n-k+۱)p}{k(۱-p)}\end{aligned}$$

بنابرین $P\{X=k\}\geq P\{X=k-۱\}$ اگر و فقط اگر $(n-k+۱)\,p\geq k\,(۱-p)$ باشد یا:

$$k\leq(n+۱)\,p$$

■

۶ متغیر تصادفی پواسون

متغیر تصادفی پواسون[۷] به عنوان تقریبی برای توزیع دو جمله‌ای با پارامترهای (n,p) وقتی که n عددی خیلی بزرگ و p عددی خیلی کوچک و np مقدار معقولی داشته باشد، به کار می‌رود. توزیع پواسون در صورتی که پارامتر λ را np تعریف کنیم به صورت زیر از روی توزیع دو جمله‌ای محاسبه می‌شود.

$$\begin{aligned}p_X(k) &= \binom{n}{k}p^k(۱-p)^{n-k}\\ &= \frac{n!}{k!(n-k)!}(\frac{\lambda}{n})^k(۱-\frac{\lambda}{n})^{n-k}\end{aligned}$$

$$=\frac{n(n-۱)...(n-k+۱)}{n^k}\times\frac{\lambda^k}{k!}\times\frac{(۱-\frac{\lambda}{n})^n}{(۱-\frac{\lambda}{n})^k}$$

حال چون n عدد خیلی بزرگی است، می‌توانیم از عبارت حد بگیریم و n را به سمت بی‌نهایت میل دهیم. داریم:

$$\lim_{n\to\infty}\frac{n(n-۱)...(n-k+۱)}{n^k}=۱$$

$$\lim_{n\to\infty}(۱-\frac{\lambda}{n})^n=e^{-\lambda}$$

$$\lim_{n\to\infty}(۱-\frac{\lambda}{n})^k=۱$$

[۷] THE POISSON RANDOM VARIABLE

بنابرین:

$$\lim_{n\to\infty} p_X(k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$$
$$k = ۰, ۱, ۲, ..$$

حال به بررسی خصوصیات این توزیع می‌پردازیم:

(آ) متوسط (امید ریاضی)

$$\begin{aligned}
\mathrm{E}[X] &= \textstyle\sum_{k=۰}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \textstyle\sum_{k=۱}^{\infty} k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \lambda \textstyle\sum_{k=۰}^{\infty} \frac{\lambda^{k-۱}}{(k-۱)!} \\
&= e^{-\lambda} \lambda e^{\lambda} \\
&= \lambda
\end{aligned}$$

(ب) واریانس

$$\begin{aligned}
\mathrm{E}[X^۲] &= \textstyle\sum_{k=۰}^{\infty} k^۲ e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \textstyle\sum_{k=۱}^{\infty} \frac{k\lambda^{k-۱}}{(k-۱)!}
\end{aligned}$$

حال به جای $k\lambda^{k-۱}$ می‌توانیم $\frac{\mathrm{d}\lambda^k}{\mathrm{d}\lambda}$ را قرار دهیم و از حسابان می‌دانیم که می‌توان جای $\sum$ و مشتق را باهم عوض کرد. پس خواهیم داشت:

$$\begin{aligned}
\mathrm{E}[X^۲] &= \lambda e^{-\lambda} \frac{\mathrm{d}}{\mathrm{d}\lambda} \textstyle\sum_{k=۱}^{\infty} \frac{\lambda^k}{(k-۱)!} \\
&= \lambda e^{-\lambda} \frac{\mathrm{d}}{\mathrm{d}\lambda} \lambda e^{\lambda} \\
&= \lambda e^{-\lambda} (e^{\lambda} + \lambda e^{\lambda})
\end{aligned}$$

بنابراین:

$$\mathrm{Var}(X) = \mathrm{E}[X^۲] - (\mathrm{E}[X])^۲ = \lambda$$

نکته ۳ *دقت کنید با توجه به اینکه توزیع پواسون تقریبی از توزیع دوجمله‌ای است، بدون محاسبات فوق نیز می‌توان به صورت حسی مقدار کمیت‌های فوق را حدس بزنیم. زیرا اگر X یک متغیر تصادفی دو جمله‌ای با پارامترهای (n, p) باشد وقتی $np \to \lambda$ و $n \to \infty$ (که نتیجه می‌دهد $p \to ۰$) داریم:*

$$\mathrm{E}[X] = np \to \lambda$$

$$\mathrm{Var}(X) = np(۱ - p) \to \lambda$$