The goal of this short note is to provide a proof and references for the "folklore fact" that Poisson random variables enjoy good concentration bounds – namely, subexponential. Thanks to Gautam Kamath for bringing the topic to my attention, and making me realize I originally had neither of the two.

Let $h\colon [-1,\infty) \to \mathbb{R}$ be the function defined by $h(u) \stackrel{\rm def}{=} 2\frac{(1+u)\ln(1+u)-u}{u^2}$.

Theorem 1. *Let $X \sim \operatorname{Poisson}(\lambda)$, for some parameter $\lambda > 0$. Then, for any $x > 0$, we have*

$$\Pr[\, X \geq \lambda + x \,] \leq e^{-\frac{x^2}{2\lambda}h\left(\frac{x}{\lambda}\right)} \tag{1}$$

and, for any $0 < x < \lambda$,

$$\Pr[\, X \leq \lambda - x \,] \leq e^{-\frac{x^2}{2\lambda}h\left(-\frac{x}{\lambda}\right)}. \tag{2}$$

In particular, this implies that $\Pr[\, X \geq \lambda + x \,], \Pr[\, X \leq \lambda - x \,] \leq e^{-\frac{x^2}{\lambda+x}}$, for $x > 0$; from which

$$\Pr[\, |X-\lambda| \geq x \,] \leq 2e^{-\frac{x^2}{2(\lambda+x)}}, \qquad x > 0. \tag{3}$$

Proof. Equations (1) and (2) are proven in Fact 5 and Fact 6, respectively. We show how they imply (3). By Fact 3, it is the case that, for every $x > 0$, $h\left(\frac{x}{\lambda}\right) \geq \frac{1}{1+\frac{x}{\lambda}}$, or equivalently $\frac{x^2}{2\lambda}h\left(\frac{x}{\lambda}\right) \geq \frac{x^2}{2(\lambda+x)}$. Thus, from (1) we get $\Pr[\, X \geq \lambda + x \,] \leq \exp(-\frac{x^2}{2\lambda}h\left(\frac{x}{\lambda}\right)) \leq \exp(-\frac{x^2}{2(\lambda+x)})$.
Similarly, for any $0 < x < \lambda$ we have $\frac{x^2}{2\lambda} > \frac{x^2}{2(\lambda+x)}$, which with (2) and Fact 2 implies $\Pr[\, X \leq \lambda - x \,] \leq \exp(-\frac{x^2}{2\lambda}h\left(-\frac{x}{\lambda}\right)) \leq \exp(-\frac{x^2}{2\lambda}h(0)) = \exp(-\frac{x^2}{2\lambda}) \leq \exp(-\frac{x^2}{2(\lambda+x)})$. ☐

Thus, we are left with proving Fact 5 and Fact 6, which we do next.

1 Establishing (1) and (2)

Fact 2. *We have $h(-1) = 2$, $h(0) = 1$, and h decreasing on $[-1,\infty)$ with $\lim_{u\to\infty} h(u) = 0$. In particular, $h \geq 0$.*

Proof. The first two properties are immediate by continuity, as, for $u \notin \{-1, 0\}$,

$$h(u) = 2\frac{(1+u)\ln(1+u)-u}{u^2} \xrightarrow[u\to-1]{} 2\frac{0-(-1)}{(-1)^2} = 2$$
$$h(u) = 2\frac{(1+u)\ln(1+u)-u}{u^2} = 2\frac{(1+u)(u-\frac{u^2}{2}+o(u^2))-u}{u^2} = 2\frac{\frac{u^2}{2}+o(u^2)}{u^2} \xrightarrow[u\to0]{} 1$$

The third property follows from differentiating the function on $(-1,0)\cup(0,\infty)$ and showing its derivative is negative; or, more cleverly, following [Pol15, Exercise 14, (ii)]. The fourth (which together with the third implies the last) directly comes from observing that $h(u) \sim_{u\to\infty} \frac{2\ln u}{u}$. ☐

Fact 3. *For any $u \geq 0$, we have $h(u) \geq \frac{1}{1+u}$.*

Proof. Consider the function $g\colon [0,\infty) \to \mathbb{R}$ defined by $g(u) = (1+u)h(u)$. We then have $g(0) = 1$, and $g(u) \sim_{u\to\infty} 2\ln u \xrightarrow[u\to\infty]{} \infty$. Moreover, by differentiation(s) (and tedious computations), one can show that g is increasing on $[0,\infty)$, which implies the claim. ☐

We follow the outline of [Pol15, Exercise 15]. For a random variable X, we denote by M its moment-generating function, i.e. $M_X : \theta \in \mathbb{R} \mapsto \mathbb{E}\left[e^{\theta X}\right]$ (provided it is well-defined). In what follows, X is a random variable following a $\operatorname{Poisson}(\lambda)$ distribution.

Fact 4. *We have $M_X(\theta) = e^{\lambda(e^\theta - 1)}$ for every $\theta \in \mathbb{R}$.*

Proof. This is a standard fact, we give the derivation for completeness. For any $\theta \in \mathbb{R}$,

$$M_X(\theta) = \mathbb{E}\big[e^{\theta X}\big] = e^{-\lambda} \sum_{n=0}^{\infty} e^{\theta n} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^\theta \lambda)^n}{n!} = e^{-\lambda} e^{e^\theta \lambda} = e^{\lambda(e^\theta - 1)}.$$

□

Fact 5. *For any $x > 0$, $\Pr[\, X \geq \lambda + x \,] \leq e^{-\frac{x^2}{2\lambda} h\left(\frac{x}{\lambda}\right)}$.*

Proof. Fix $x > 0$. For any $\theta \in \mathbb{R}$,

$$\Pr[\, X \geq \lambda + x \,] = \Pr\left[e^{\theta X} \geq e^{\theta(\lambda + x)} \right] = \Pr\left[e^{\theta(X - \lambda - x)} \geq 1 \right] \leq \mathbb{E}\left[e^{\theta(X - \lambda - x)} \right]$$

recalling that if Y is a discrete random variable taking values in $\mathbb{N}$, $\Pr[\, Y > 0 \,] = \Pr[\, Y \geq 1 \,] = \sum_{n=1}^{\infty} \Pr[\, Y = n \,] \leq \sum_{n=1}^{\infty} n \Pr[\, Y = n \,] = \mathbb{E}[Y]$. Rearranging the terms and taking the infimum over all $\theta > 0$, we have

$$\begin{aligned} \Pr[\, X \geq \lambda + x \,] &\leq \inf_{\theta > 0} \mathbb{E}\big[e^{\theta X}\big]\, e^{-\theta(\lambda + x)} = \inf_{\theta > 0} e^{\lambda(e^\theta - 1)} e^{-\theta(\lambda + x)} && \text{(Fact 4)} \\ &= \inf_{\theta > 0} e^{\lambda(e^\theta - 1) - \theta(\lambda + x)} = e^{\inf_{\theta > 0}(\lambda(e^\theta - 1) - \theta(\lambda + x))}. \end{aligned}$$

It is a simple matter of calculus to find that $\inf_{\theta > 0}(\lambda(e^\theta - 1) - \theta(\lambda + x))$ is attained for $\theta^* \stackrel{\text{def}}{=} \ln(1 + \frac{x}{\lambda}) > 0$, from which

$$\Pr[\, X \geq \lambda + x \,] \leq e^{\lambda(e^{\theta^*} - 1) - \theta^*(\lambda + x)} = e^{-\lambda((1 + \frac{x}{\lambda}) \ln(1 + \frac{x}{\lambda}) - \frac{x}{\lambda})} = e^{-\frac{x^2}{2\lambda} h\left(\frac{x}{\lambda}\right)}$$

as claimed. □

Fact 6. *For any $0 < x < \lambda$, $\Pr[\, X \leq \lambda - x \,] \leq e^{-\frac{x^2}{2\lambda} h\left(-\frac{x}{\lambda}\right)} \leq e^{-\frac{x^2}{2\lambda}}$.*

Proof. Fix $0 < x < \lambda$. As before, for any $\theta \in \mathbb{R}$,

$$\Pr[\, X \leq \lambda - x \,] = \Pr\left[e^{\theta X} \leq e^{\theta(\lambda - x)} \right] = \Pr\left[e^{\theta(\lambda - x - X)} \geq 1 \right] \leq \mathbb{E}\big[e^{-\theta X}\big]\, e^{\theta(\lambda - x)}.$$

Rearranging the terms and taking the infimum over all $\theta > 0$, we have

$$\begin{aligned} \Pr[\, X \leq \lambda - x \,] &\leq \inf_{\theta > 0} \mathbb{E}\big[e^{-\theta X}\big]\, e^{\theta(\lambda - x)} = \inf_{\theta > 0} e^{\lambda(e^{-\theta} - 1)} e^{\theta(\lambda - x)} && \text{(Fact 4)} \\ &= e^{\inf_{\theta > 0}(\lambda(e^{-\theta} - 1) + \theta(\lambda - x))}. \end{aligned}$$

It is again straightforward to check, e.g. by differentiation, that $\inf_{\theta > 0}(\lambda(e^{-\theta} - 1) + \theta(\lambda - x))$ is attained for $\theta^* \stackrel{\text{def}}{=} -\ln(1 - \frac{x}{\lambda}) > 0$, from which

$$\Pr[\, X \leq \lambda - x \,] \leq e^{\lambda(e^{-\theta^*} - 1) + \theta^*(\lambda - x)} = e^{-x - (\lambda - x) \ln(1 - \frac{x}{\lambda})} = e^{-\lambda((1 - \frac{x}{\lambda}) \ln(1 - \frac{x}{\lambda}) + \frac{x}{\lambda})} = e^{-\frac{x^2}{2\lambda} h\left(-\frac{x}{\lambda}\right)}$$

as claimed. The last step is to observe that, by Fact 2, $e^{-\frac{x^2}{2\lambda} h\left(-\frac{x}{\lambda}\right)} \leq e^{-\frac{x^2}{2\lambda} h(0)} = e^{-\frac{x^2}{2\lambda}}$. □

2 An alternative proof of (1)

Recall that if $(Y^{(n)})_{n\geq 1}$ is a sequence of independent random variables such that $Y^{(n)}$ follows a $\mathrm{Bin}\big(n, \frac{\lambda}{n}\big)$ distribution, then $(Y^{(n)})_{n\geq 1}$ converges in law to X, a random variable with Poisson(λ) distribution.[1] In particular, since convergence in law corresponds to pointwise convergence of distribution functions, this implies that, for any $t \in \mathbb{R}$,

$$\Pr\Big[\, Y^{(n)} \geq t \,\Big] \xrightarrow[n\to\infty]{} \Pr[\, X \geq t \,]\,. \tag{4}$$

For any fixed $n \geq 1$, we can by definition write $Y^{(n)}$ as $Y^{(n)} = \sum_{k=1}^{n} Y_k^{(n)}$, where $Y_1^{(n)}, \dots, Y_n^{(n)}$ are i.i.d. random variables with $\mathrm{Bern}\big(\frac{\lambda}{n}\big)$ distribution. Note that $\mathbb{E}\big[Y^{(n)}\big] = \lambda$ and $\mathrm{Var}[Y^{(n)}] = \lambda(1 - \frac{\lambda}{n}) \leq \lambda$. As $\mathbb{E}\Big[Y_k^{(n)}\Big] = \frac{\lambda}{n}$ and $|Y_k^{(n)}| \leq 1$ for all $1 \leq k \leq n$, we can apply Bennett's inequality ([BLM13, Chapter 2],[Pol15, Chapter 2.5]), to obtain, for any $t \geq 0$,

$$\Pr\Big[\, Y^{(n)} \geq \lambda + x \,\Big] = \Pr\Big[\, Y^{(n)} \geq \mathbb{E}\Big[Y^{(n)}\Big] + x \,\Big] \leq e^{-\frac{x^2}{2\lambda} h\left(\frac{x}{\lambda}\right)}$$

Taking the limit as n goes to ∞, we obtain by (4) that $\Pr[\, X \geq \lambda + x \,] \leq e^{-\frac{x^2}{2\lambda} h\left(\frac{x}{\lambda}\right)}$, re-establishing (1).

Remark 7. We note that a qualitatively similar statement (yet quantitatively weaker) can be obtained by observing that Poisson distributions are in particular (discrete) log-concave, and that any log-concave (discrete or continuous) has subexponential tail [An95].

Remark 8. As another way to establish the result, we refer the reader to [Gol17, Proposition 11.15], where bounds on individual summands of the Poisson tails are obtained. From there, one can attempt to derive Theorem 1, specifically (3).

References

[An95] M. Y. An. Log-concave probability distributions: Theory and statistical testing. Technical Report Economics Working Paper Archive at WUSTL, Washington University at St. Louis, 1995. 7

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013. 2

[Gol17] Oded Goldreich. *Introduction to Property Testing*. Forthcoming, 2017. Preliminary version accessible at `http://www.wisdom.weizmann.ac.il/~oded/pt-intro.html` (accessed 02-23-2017). 8

[Pol15] David Pollard. MiniEmpirical. `http://www.stat.yale.edu/~pollard/Books/Mini/`, 2015. Manuscript (accessed 02-23-2017). 1, 1, 2, 0

[0] This approach is inspired by [Pol15, Exercise 16]).